



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :	A1	(11) International Publication Number: WO 98/42738
C07K 14/00		(43) International Publication Date: 1 October 1998 (01.10.98)
(21) International Application Number: PCT/US98/05311		60/056,370 19 August 1997 (19.08.97) 60/060,862 2 October 1997 (02.10.97) US
(22) International Filing Date: 19 March 1998 (19.03.98)		
(30) Priority Data:		(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).
60/041,281 21 March 1997 (21.03.97) US		(72) Inventors; and
60/041,276 21 March 1997 (21.03.97) US		(75) Inventors/Applicants (for US only): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US).
60/042,344 21 March 1997 (21.03.97) US		GREENE, John, M. [US/US]; 872 Diamond Avenue, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US).
60/041,277 21 March 1997 (21.03.97) US		RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US).
60/048,355 30 May 1997 (30.05.97) US		DUAN, Roxanne [US/US]; 4541 Fairfield Drive, Bethesda, MD 20814 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US).
60/048,096 30 May 1997 (30.05.97) US		EBNER, Reinhart [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 M. Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US).
60/048,351 30 May 1997 (30.05.97) US		
60/048,154 30 May 1997 (30.05.97) US		
60/048,160 30 May 1997 (30.05.97) US		
60/048,069 30 May 1997 (30.05.97) US		
60/048,131 30 May 1997 (30.05.97) US		
60/048,186 30 May 1997 (30.05.97) US		
60/048,095 30 May 1997 (30.05.97) US		
60/048,187 30 May 1997 (30.05.97) US		
60/048,099 30 May 1997 (30.05.97) US		
60/050,937 30 May 1997 (30.05.97) US		
60/048,352 30 May 1997 (30.05.97) US		
60/048,135 30 May 1997 (30.05.97) US		
60/048,188 30 May 1997 (30.05.97) US		
60/048,094 30 May 1997 (30.05.97) US		
60/048,350 30 May 1997 (30.05.97) US		
60/054,804 5 August 1997 (05.08.97) US		
(74) Agents: BROOKES, Anders, A. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US).		
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).		
(54) Title: 87 HUMAN SECRETED PROTEINS		Published With international search report.

(57) Abstract

The present invention relates to 87 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## 87 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or 10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum 15 (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

20 Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or 25 secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes 30 encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying 35 and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, 5 and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

10

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original 15 environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

20 In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce 25 a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid 30 sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated 35 amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, 5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained 10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the 15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages 20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even 25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include 30 Denhardt's reagent, BLOTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such 35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5       The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and 10 double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability 15 or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

20       The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, 25 as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or 30 without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, 35 covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins  
5 such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990);  
10 Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y"  
refers to a polypeptide sequence, both sequences identified by an integer specified in  
Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting  
15 activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present  
20 invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

25 **Polynucleotides and Polypeptides of the Invention**

**FEATURES OF PROTEIN ENCODED BY GENE NO: 1**

The translation product of this gene shares sequence homology with nucleolin, which is thought to be important in macromolecule binding, as well as some membrane  
30 proteins. Preferred polypeptide fragments comprise the amino acid sequence:  
DPEAADSGEQNKRTPDLPEEEYVKEEIQENEEAVKKMLVEATREFEEVVVDES  
(SEQ ID NO:239); QKLKRKAEDPEAADSGEQNKRTPDLPEEEYVKEEIQENEE  
AVKKMLVEATREFEEVVVDES (SEQ ID NO:240); KAMEKSSLTQHSWQSLKDR  
YLKHRLRGQEHKYLLGDAPVSPSSQKLKRKAEDPEAADSGEQNKRTPDLPEE  
35 EYVKEEIQENEEAVKKMLVEATREFEEVVVDESPPDFEIHI (SEQ ID NO:241).  
Also preferred are the polynucleotide fragments encoding these polypeptide fragments.

This gene maps to chromosome 16, and therefore can be used as a marker in linkage analysis for chromosome 16.

This gene is expressed primarily in brain and kidney and to a lesser extent in wide range of tissues.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cell-cell interaction or cell-matrix interaction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes  
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and kidney, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,  
15 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:125 as residues: Met-1 to Trp-10.

The tissue distribution and homology to nucleolin indicates that polynucleotides  
20 and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases involving cell-cell interaction or cell-extracellular matrix interaction.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 2

The translation product of this gene shares sequence homology with a porcine  
25 zona pellucida protein ZPDS.1711. (See Accession No. R39356.) These two proteins have weak homology with *Drosophila* commissureless and metal homeostasis proteins which are thought to be important in controlling growth cone guidance across the CNS midline and protecting cells against reactive oxygen toxicity. thus, based on homology, it is likely that this gene also be involved in development. Preferred polypeptide  
30 fragments comprise the amino acid sequence: LPSYDEAERTKAEATIPLVPGRDEDVF  
VGRDDFDDADQLRIGNDGIFMLTFFMAFLFNWIGFFLSFCLTTSAAGRYGAISG  
FGSLIKWILIVRFSTYFPGYFDGQYWLVWVFLVLGFLLFLRGFINYAKVRKM  
PETFSNLPRTRVLFI (SEQ ID NO:242); and/or AGRYGAISGFGLSLIKWILIVRFS  
(SEQ ID NO:243). Also preferred are polynucleotide fragments encoding these  
35 polypeptide fragments. This gene maps to chromosome 5, and therefore can be used in linkage analysis as a marker for chromosome 5.

This gene is expressed primarily in kidney, adrenal gland, brain and to a lesser extent in wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, fertilization control or tissue damages by metabolites or other toxic agents. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and urosecretion system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, adrenal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zona pellucida protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for fertility control such as contraceptive development. The homology with metal homeostasis and commissureless genes indicates the gene's function in spermatozoa guidance and protection. It would also be useful for the treatment/diagnosis of tissue damages caused by toxic metabolites and other agents since the gene product is also expressed in urosecretive tissues.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in liver and to a lesser extent in placenta. Preferred polypeptide fragments comprise the amino acid sequence: MKHLSAWNFT KLTFLQLWEI FEGSVENCQTLTYSKLIQIKYTFSRGSTFYI (SEQ ID NO:244). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, digestive and nutrient transport/utilization disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and

circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., liver, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, 5 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in liver and placenta indicates that the protein product is either an extracellular enzyme or a molecule carrier. Therefore, polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/treatment of digestive 10 and nutrient transport/utilization disorders, including malabsorption and malnutrition.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene shares homology with the *sap47* gene of *Drosophila melanogaster*, a gene which codes for a conserved neuronal protein associated with synaptic terminals. (See 15 Mol. Brain Res. 32:45-54 (1995); see also, Accession No. 929571.) Thus, based on homology, the gene of the present invention also should be associated with synaptic terminals. Preferred polypeptide fragments comprise the amino acid sequence:  
FSSDFRTSPWESRRVESKATSARCGLWGSGPRRRPASGMFRGLSSWLGLQQP  
VAGGGQPNGDAPPEQPSETVAESAEEELQQAGDQELLHQAKDFGNYLFNFASA  
20 ATKKITESVAETAQTIKKVEEGKIDGIIDKTIIGDFQKEQKKFVEEQHTKKSEA  
AVPPWVDTNDEETIQQQILALSADKRNLRDPPAGVQFNFDQMDQMPVALVML  
(SEQ ID NO:245); MRFALVPKLVKEEVFWRNYFYRVSLIKQSAQLTALAAQQQA  
AGKGGEQQ (SEQ ID NO:246); STSPGVSEFVSDAFDACNLNQEDLRKEMEQL  
VLDKKQEETAVLEEDSADWEKELQQELQEYEVVTESEKRDENWDK (SEQ ID  
25 NO:247); SPWESRRVESKATSARCGLWGSGPRRRPASGMFRGLSSWLGLQQ  
PVAGGGQPNGDAPPEQPS (SEQ ID NO:248); PVAGGGQPNGDAPPEQPSETV  
ESAEEELQQAGDQELLHQAKDFGNYLFNFASAATKKITESVAE (SEQ ID NO:  
249); and/or FQKEQKKFVEEQHTKKSEAAVPPWVDTNDEETIQQQILALSADKR  
NFLRDPPAGVQFNFDQMDQMPVALVML (SEQ ID NO:250). Also preferred are  
30 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in kidney pyramids and to a lesser extent in lung and other tissues of various types. This gene fluxes calcium in human aortic smooth muscle cells, and therefore is involved in signal transduction.

Therefore, polynucleotides and polypeptides of the invention are useful as 35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, renal and nervous disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and/or nervous system, expression 5 of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, lung, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in 10 healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in kidney and lung and homology with sap47 indicates that the protein product has regulatory or direct functions in molecular exchange with body fluids and nervous system signaling. Polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in kidney and nervous 15 system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 5

The translation product of this gene shares sequence homology with the mouse Ly-9.2 antigen which is thought to be an important cell surface marker in lymphoids, 20 myeloids and hematopoietic progenitors. (See Accession No. gil198932.) Preferred polypeptide fragments comprise the amino acid sequence: PFICVARNPVSRNFSSPI LARKLCEGAA (SEQ ID NO:251); and/or KEDPANTVYSTVEIPKKMENPHSLLT MPDTPRL (SEQ ID NO:252). Also preferred are polynucleotide fragments encoding 25 these polypeptide fragments. Based on homology, it is likely that this gene is also a cell surface marker, involved in hematopoiesis.

This gene is expressed primarily in activated macrophages, monocytes and T-cells and to a lesser extent in spleen and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, 35 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., blood cells, and bone marrow, and cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those 5 comprising a sequence shown in SEQ ID NO:129 as residues: Lys-26 to Tyr-33, Arg-44 to Ile-49, Ser-53 to Lys-71, Lys-86 to Pro-91.

The tissue distribution and homology to Ly-9.2 surface immunoglobulin family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune and hematopoietic disorders. Polypeptides and polynucleotides 10 corresponding to this gene are also be used as a marker for leukemia or a modulator of the functions of the cells of macrophage/monocyte or T-cell types.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with the 15 *Drosophila* glutactin gene which is thought to be important in cell-cell interaction or cell-extracellular matrix contact.

This gene is expressed primarily in colon tissue, aorta endothelial cells and to a lesser extent in skin, breast tissue and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as 20 reagents for differential identification of these tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the gastrointestinal tract, vascular system or T-cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these 25 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, cardiovascular system, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., colon, cardiovascular tissue, skin, mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, 30 synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to glutactin indicates that polynucleotides 35 and polypeptides corresponding to this gene are useful for the development and maintenance of the integrity of the basal membrane in the gastrointestinal tract and

cardiovascular system. The expression in T-cells also indicate the protein may be involved in T-cell adhesion, cell-cell interaction and development.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 7

5 The translation product of this gene shares sequence homology with MURF4 protein, an ATPase homolog, which is thought to be important in ATP hydrolysis.

This gene is expressed primarily in breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 10 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer and non-neoplastic breast diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast tissue, expression of 15 this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an 20 individual not having the disorder.

The tissue distribution and homology to MURF4 gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neoplastic or non-neoplastic breast diseases because ATPase like protein may be involved in changed metabolic states of the breast.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene shares homology to the alcohol dehydrogenase gene. Preferred polypeptide fragments comprise the amino acid sequence: ASA VLL DLPNSG GEA QAK KLG NNC VFA PADV TSE KDV QTAL ALA KGK FGR DV A VNC AGIAV AS 30 KT YNL KKQ QTHT LED FQR VL DV NLM GTF NVIR LVAG EMG QNE PDQ GG QRG VI INT ASV AAF EGQ VGQ AA Y SASK GGIV GM TLPI ARDL API GIV VMTI APGL FG TP L TS LPE KVC NFL ASQ VP FP SRL GDPA EY AHL VQAI IENP FLN GEV IR LDG AIR MQ P (SEQ ID NO:253); and/or SVA AF EGQ VGQ AA Y SASK GGIV GM TLPIA (SEQ ID NO:254). Polynucleotides encoding these fragement are also encompassed by the 35 invention. Other groups have also recently cloned this gene, recognizing its homology to alcohol dehydrogenase. (See Accession No. 1778355.) Moreover, a second group

recently cloned the mouse homologue of this gene. (See Accession No. 2078284.) They found that the mouse homologue binds to amyloid beta-peptide and mediates neurotoxicity in Alzheimer's disease, calling the protein ERAB. This gene maps to chromosome X, and therefore can be used in linkage analysis as a marker for  
5 chromosome X. Therefore, mutations in the translated product of this gene may be involved in Alzheimer's disease in humans, as well as other sex linked diseases. This gene can be used as a diagnostic marker for these diseases.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:132 as residues: Arg-45 to Ser-53.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

The translation product of this gene shares weak sequence homology with rat N-methyl-D-aspartate receptor subunit and other proline-rich proteins which are thought to be important in neurotransmission or protein-protein interaction.

15 This gene is expressed primarily in synovial hypoxia and to a lesser extent in ovary, senescent cells and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are  
20 not limited to, synovial hypoxia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the synovia and brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
25 synovial tissue, ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the  
30 disorder.

The tissue distribution in synovial hypoxia and nerve tissues, and homology to N-methyl-D-aspartate receptor subunit and other proline-rich proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of synovial hypoxia and other synovial disorders.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 10**

This gene is expressed primarily in prostate and to a lesser extent in placenta and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male and female infertility, cancer, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., prostate, placenta, ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:134 as residues: Pro-17 to Met-23, Ala-30 to Trp-38, Ile-49 to Trp-54, Lys-68 to Gly-74, Thr-93 to Gly-99, Met-126 to Glu-132, Gly-173 to Ser-178, Lys-205 to Tyr-214.

The tissue distribution of this gene in the prostate, placenta and ovary indicates that this gene product is useful for treatment/diagnosis of male or female infertility, endocrine disorders, fetal deficiencies, ovarian failure, amenorrhea, ovarian cancer, benign prostate hyperplasia, prostate cancer, and other forms of cancer of the reproductive system.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 11**

This gene is expressed primarily in the thyroid and to a lesser extent in the pineal gland. This gene maps to chromosome 10, and therefore can be used as a marker in linkage analysis for chromosome 10. Preferred polypeptide fragments comprise the amino acid sequence: HPIEWAINAATLSQFY (SEQ ID NO:256); CWIKYCLTLMQN AQLSMQDNIG (SEQ ID NO:257); KVSYLRPLDFEEARELFLLGQHYVF (SEQ ID NO:258); MERRCKMHKRXIAMEPLTVDLNPQ (SEQ ID NO:259); and/or SHIV KKINNLNKSALKY YQLFLD (SEQ ID NO:260). Also preferred are polynucleotides encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, thyroid and pineal gland disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
- 5       for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., thyroid and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another
- 10      tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:135 as residues: Ser-2 to Ser-8, Thr-38 to Arg-44.
- 15      The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/detecting immune disorders such as arthritis, asthma, immune deficiency diseases (e.g., AIDS), and leukemia, as well as treating/detecting thymus disorders (e.g., Graves Disease, lymphocytic thyroiditis, hyperthyroidism, and hypothyroidism), and treating/detecting pineal gland disorders
- 20      (e.g., circadian rhythm disturbances associated with shift work, jet lag, blindness, insomnia and old age).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 12

- This gene is expressed primarily in lung and tonsils.
- 25      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary or immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for
- 30      differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pulmonary tissue, and tonsils, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
- 35      another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:136 as residues: Glu-28 to Gly-49.

The tissue distribution of this gene only in lung indicates that it could play a role in the treatment/detection of lung lymphoma or sarcoma formation, pulmonary edema and embolism, bronchitis and cystic fibrosis. Its expression in tonsils indicates a potential role in the treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases (e.g., AIDS), and leukemia, in addition to the treatment/detection of tonsillitis.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in lymphoid, myeloid and erythroid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, myeloid cells, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The predominant tissue distribution of this gene in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma, immunodeficiency diseases and leukemia. Preferred embodiments of the present invention are polypeptide fragments comprising the amino acid sequence: FTHLSTCLLSLLLVRMSGFLLARASPSI  
30 CALDSSCFVEYCSSYSSSCFLHQHFPSLLDHLCQ (SEQ ID NO:261); or FLLL ARASPSICALDSSCFVQEY (SEQ ID NO:262). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 14

35 This gene is homologous to the *Drosophila Regena* (*Rga*) gene. (See Accession No. 1658504.) This *Drosophila* gene is thought to be a homolog of the global negative

transcriptional regulator NOT2 (CDC36) from yeast, which modifies gene expression and suppresses position effect variegation. Preferred polypeptide fragments comprise the amino acid sequence: PDGRVTNIPQGMVTDQFGMIGLLTFIRAAETDPGMVHL ALGSDLTTLGLNLNS (SEQ ID NO:263); VHLALGSDLTTLGLNLNSPENLYP (SEQ ID NO:265); EDLLFYLYYMNGGDVLQLLAAVELFNRDWRYHKEERVWI TR (SEQ ID NO:264); and/or HNEDFPALPGS (SEQ ID NO:266).

5 This gene is expressed primarily in placenta and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of  
15 disorders of the above tissues or cells, particularly of the neurological system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
20 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:138 as residues: Leu-9 to Tyr-15, Asp-34 to Gln-46, Pro-51 to Asp-57, Gly-88 to Thr-104, Thr-123 to Ser-128.

25 The tissue distribution of this gene indicates that it could be used in the detection and/or treatment of neurological disorders such as such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, and panic disorder.

### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 15

This gene is expressed primarily in adrenal gland tumor and osteoclastoma.  
Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and bone disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system and in bone, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adrenal gland, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:139 as residues: Ile-52 to Trp-57.

The tissue distribution of this gene indicates that it may be involved in the treatment and/or detection of adrenal gland tumors, osteosarcomas, endocrine disorders and bone disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with the FK506 binding protein, a protein which plays an important role in immunosuppression. (See Accession No. M75099.) Specifically, a 12-kDa FK506-binding protein (FKBP-12) is a cytosolic receptor for the immunosuppressants FK506 and rapamycin. (See, Proc. Natl. Acad. Sci. 88: 6677-6681 (1991).) Thus, based on homology, it is likely that this gene also has immunosuppression activity. Preferred polypeptides comprise the amino acid sequence: GRIIDTSLTRDPLVIELGQKQVIPGLEQSLLDMCVGEKRRAIIPSH LAYGKRGFPPSVPADAVVQYDVELIALIR (SEQ ID NO:267); and/or IHYTGSILV DGR IIDTS (SEQ ID NO:268). Also preferred are the polynucleotide fragments encoding these polypeptides.

This gene is expressed primarily in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:140 as residues: Ala-118 to Phe-124, Arg-178 to Lys-201.

5       The tissue distribution and homology to the FK506 binding proteins which are believed to a role in immunosuppression mediated by the immunosuppressant drugs rapamycin and cyclosporin, indicates that this gene could serve as a novel target for the identification of novel immunosuppressant drugs.

10      **FEATURES OF PROTEIN ENCODED BY GENE NO: 17**

The translation product of this gene shares sequence homology with the rat calcium-activated potassium channel rSK3, which is thought to be important in regulating vascular tone. (See Accession No. gil2564072, gil1575663, and gil1575661.) Although homologous to these proteins, this gene contains an 18 amino acid insert, not previously identified in the homologs. Preferred polypeptide fragments comprise the amino acid sequence: CESPESPAQPSGSSLPAWYH (SEQ ID NO:269). Also preferred are the polynucleotide fragments encoding these polypeptides.

This gene is expressed primarily in B-cells, frontal cortex and endothelial cells. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular (hyper/hypotension, asthma, pulmonary edema, pneumonia, heart disease, restenosis, atherosclerosis, stroke, angina and thrombosis) and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, and endothelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:141 as residues: Glu-72 to Gly-82, His-90 to Val-95, Gln-168 to Lys-174, Val-202 to Ser-212.

5       The tissue distribution and homology to calcium-activated potassium channels indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of vascular disorders (hyper/hypotension, athesma, pulmonary edema, pneumonia, heart disease, restenosis, atherosclerosis, stoke, angina and thrombosis).

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 18**

This gene is expressed primarily in smooth muscle and to a lesser extent in brain (amygdala, corpus colosum, hippocampus).

10     Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular (hypertension, heart disease, athesma, pulmonary edema, restenosis, atherosclerosis, stoke, angina, thrombosis, and wound healing), and  
15     neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
20     smooth muscle, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include  
25     those comprising a sequence shown in SEQ ID NO:142 as residues: Lys-43 to Arg-49, Tyr-58 to Glu-65.

30     The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (hypertension, heart disease, athesma, pulmonary edema, restenosis, atherosclerosis, stoke, angina, thrombosis, and wound healing). Expression in brain indicates a role in the treatment and diagnosis of behavioral or neurological disorders, such as depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior.

35      **FEATURES OF PROTEIN ENCODED BY GENE NO: 19**

This gene is expressed primarily in T-cells (Jurkats, resting, activated, and

anergic T-cells), endothelial cells, pineal gland, and to a lesser extent in a variety of other tissues and cell types. Preferred polypeptide fragments comprise the amino acid sequence: EEAGAGRRCSHGGARPAGLGNEGLGLGGDPDHTDTGSRSKQRINN WKE SKHKVIMASASARGNQDKDAHFPPPSKQSLLFCPKSKLHIHRAEISK  
5 (SEQ ID NO:270); and/or SKQRINNWKE SKHKVIMASASAR (SEQ ID NO:271). Also preferred are the polynucleotide fragments encoding these polypeptides.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, immune and cardiovascular disorders. Similarly, 10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neurological and vascular systems, expression of this gene at significantly higher or 15 lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, endothelial cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily 20 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:143 as residues: Phe-71 to Arg-76, Pro-82 to His-87, Glu-103 to Ala-111.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune 25 disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. In addition, expression in the pineal gland might suggest a role in the diagnosis of specific brain tumors and treatment of neurological disorders. Endothelial cell expression might suggest a role in cardiovascular or respiratory/pulmonary disorders or infections 30 (athesma, pulmonary edema, pneumonia).

## FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in brain and embryo and to a lesser extent in leukocytes. This gene maps to chromosome 15, and therefore can be used as a marker 35 in linkage analysis to chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes 5 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from 10 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:144 as residues: Met-1 to Gly-8.

The tissue distribution indicates that polynucleotides and polypeptides 15 corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain -- and in particular the fetal brain -- would suggest a possible role in the treatment and diagnosis of developmental and neurodegenerative diseases of 20 the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in brain, kidney, lung, liver, spleen, and a 25 variety of leukocytes (especially T-cells) and to a lesser extent in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemias, lymphomas, autoimmune, immunosuppressive, and 30 immunodeficiencies, hematopoietic disorders, as well as renal disorders, and neoplasms. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal, pulmonary, immune, and central nervous systems, expression of this gene at 35 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

- brain and other tissue of the nervous system, kidney, pulmonary tissue, liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
- 5 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of renal conditions, such as acute renal failure, kidney fibrosis, and kidney tubule regeneration.

10 The expression in leukocytes and other immune tissues indicates a role in immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain -- and in particular the fetal brain -- indicates a possible role in the treatment and diagnosis of developmental and neurodegenerative diseases of the

15 brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in skin (fetal epithelium, keratinocytes and skin). This gene also maps to chromosome 19, and therefore can be used in linkage analysis as a marker for chromosome 19.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skin cancers (e.g., melanomas), eczema, psoriasis or other disorders of the skin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., skin and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:146 as residues: Pro-28 to Glu-35, Ser-39 to Phe-44, Ala-94 to Gln-99.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of skin cancers (e.g., melanomas), eczema, psoriasis or other disorders of the skin.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 23

This gene maps to chromosome 11. Another group recently isolated this same gene, associating the sequence to the region thought to harbor the gene involved in Multiple Endocrine Neoplasia Type 1, or MEN 1. (See Accession No. 2529721 and Genome Res. 7(7), 725-735 (1997), incorporated herein by reference in its entirety.)

10 Preferred polypeptide fragments comprise the amino acid sequence: LFHWACLNERA AQLPRNTAXAGYQCPSCNGPS (SEQ ID NO:272).

This gene is expressed primarily in epididymus, pineal gland, T-cells, as well as fetal epithelium, lung and kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, metabolic mediated disorders, and MEN. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, renal, neurological and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epididymus and other reproductive tissue, pineal gland, T-cells and other blood cells, epithelium, lung, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of developmental deficiencies or abnormalities as well as a host of different disorders which arise as a result of conditions in the indicated tissues or cell types. An area of particular interest is in the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain, and in particular the fetal brain, would suggest a possible role in the treatment and diagnosis of

- developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior). Respiratory/pulmonary disorders, such as athesma, pulmonary edema are also potential therapeutic areas, as well as renal conditions such as acute renal failure, kidney fibrosis and kidney tubule regeneration. Moreover, this gene can be used in the treatment and/or detection of MEN I.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in fetal spleen.

- 10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia, lymphoma, AIDS, hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing 15 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., spleen and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or 20 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune 25 disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

- A closely related homolog of this gene was recently cloned by another group, 30 calling the gene CDO, an oncogene-, serum-, and anchorage-regulated member of the Ig/fibronectin type III repeat family. (See Accession No. 2406628, and J. Cell Biol. 138(1): 203-213 (1997), herein incorporated by reference in its entirety.) Preferred polypeptide fragments comprise the amino acid sequence: FYTYYRPTDSDNDSDYKK DMVEGDKYWHSISHLQPETSYDIKMQCFNEGGESEFSNVMICETKARKSSGQP 35 GRLPPPTLAPPQPLPETIERPVGTGAMVARSSDLPYLIVGVVLGSIVLIIVTFIPF CLWRAWSKQKHTTDLGFPRLALPPSCPYTMVPLGLPGHQAVDSPTSVASVD

GPVLM (SEQ ID NO:273); or YIYYRPTSDNDSDYKKDMVEGDKYWHSISHLQ  
PETSYDIKMQCFNEGGESEFSNVMICETKARKS (SEQ ID NO:274).

This gene is expressed primarily in fetal lung and kidney, human embryo and osteoclastoma stromal cells and to a lesser extent in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders and cancers, as well as pulmonary and renal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory/pulmonary, skeletal and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, kidney, embryonic tissue, and bone cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:149 as residues: Thr-5 to Pro-18, Ala-76 to Thr-84.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of: osteoperosis, fracture, osteosarcoma, ossification, and osteonecrosis, as well as respiratory/pulmonary disorders, such as athesma, pulmonary edema, and renal conditions such as acute renal failure, kidney fibrosis and kidney tubule regeneration.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 26

This gene is homologous to the HIV envelope glycoprotein. (See Accession No. 2641463.) Preferred polypeptide fragments comprise the amino acid sequence: NVRALLHRMPEPPKINTAKFNNNKRKNLSL (SEQ ID NO:275).

This gene is expressed primarily in pineal gland and skin, and to a lesser extent in lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

- not limited to, neurological and behavior disorders; respiratory/pulmonary disorders, such as athesma, pulmonary edema; skin conditions such as eczema, psoriasis, acne and skin cancer, as well as AIDS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and respiratory systems, as well as skin and AIDS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, pineal gland, epidermis, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:150 as residues: Gln-15 to Gln-20.
- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions which affect the above tissues, such as: skin cancer, eczema, psoriasis, acne, athesma, pulmonary edema, neuro-degenerative or developmental disorders such as Alzheimer's, depression, schizophrenia, dementia, and AIDS.

20 **FEATURES OF PROTEIN ENCODED BY GENE NO: 27**

Preferred polypeptide encoded by this gene comprise the following amino acid sequence: NTNQREALQYAKNFQPFALNHQKDIQVLMGSLVYLRQGIENSPYVHL LDANQWADICDIFTRDACALLGLSVESPLSVSAGCVALPALNIKAVIEQRQC 25 TGVWNQKDELPIEVDLGKKCWYHSIFACPILRQQTTDNNPPMKLVCGHIIISRD ALNKMFDNGSKLKCPYCPMEQSPGDAKQIFF (SEQ ID NO:276). Polynucleotides encoding such polypeptides are also provided as are complementary polynucleotides thereto.

This gene is expressed primarily in liver (adult and fetal) and spleen tissue, and 30 to a lesser extent in placenta, T helper cells, kidney tumor, ovarian tumor, melanocytes and fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 35 not limited to, immune and developmental diseases and disorders and liver diseases such as liver cancer. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, circulatory and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues  
5 and cell types (e.g., liver, spleen, placenta, blood cells, kidney, ovary and other reproductive tissue, melanocytes, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
10 fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of growth, hematopoietic and immune system disorders particularly related to the liver.

## 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with prostaglandin transporter which is thought to be important in metabolic and endocrine disorders. See, for example, Gastroenterology Oct:109(4):1274-1282 (1995). Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

20 SYLSACFAGCNSTNLGCACTTVPAENATVVPGKCPSPGCQEAFLTFLCVMCI  
CSLIGAMARHP (SEQ ID NO:277); and/or PSVIIILRTVSPELKSYALGVLFLLRL  
LGFIPPLIFGAGIDSTCLFWSTFCGEQGACVLYDNVVYRYLYVSIAIALKSFAFI  
(SEQ ID NO:278).

This gene is expressed primarily in hematopoietic and brain tissues.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic, immune and endocrine diseases and disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing  
30 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., endocrine tissue, hematopoietic tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to  
35

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to prostaglandin (and anion) transporter indicates that polynucleotides and polypeptides corresponding to this gene are useful for 5 study, diagnosis and treatment of endocrine, metabolic, immune and kidney disorders.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 29**

This gene is expressed primarily in early stage human lung.

Therefore, polynucleotides and polypeptides of the invention are useful as 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of 15 the above tissues or cells, particularly of the developmental and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the 20 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:153 as residues: Val-50 to Trp-55.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of respiratory and growth diseases and disorders.

25

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 30**

The translation product of this gene shares sequence homology with human DNA helicase which is thought to be important in accurate and complete DNA replication in creation of new cells. Preferred polypeptides encoded by this gene 30 comprise the following amino acid sequence: QSLFTRFVRGVGVPTVLDDAQGRARA SLCXXYNWRYKNLGNLPHVQLLPEFSTANAGLLYDFQLINVEDFQGVGESEPN PYFYQLGEAEYVVALFMYMCLLGYPADKISILTTYNGQKHLIRDINRRCGNN PLIGRPNKVTVDRFQGQQNDYILLSLVRTRAVGHLRDVRRLVVAMSRAR (SEQ ID NO:279); and/or LVKEAKIIAMTCTHAALKRHDVKLGFKYDNILMEE 35 AAQILEIETFIPLLLQNPQDGFSRLKRWIMIGDHQLPPVI (SEQ ID NO:280).

This gene is expressed primarily in testes tumor and to a lesser extent in adrenal

gland tumor and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 5 not limited to, cancers and endocrine/growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine, developmental, and reproductive systems, expression of this gene at significantly higher or lower levels 10 may be routinely detected in certain tissues (e.g., testes and other reproductive tissue, adrenal gland, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an 15 individual not having the disorder.

The tissue distribution and homology to DNA helicase indicates that the protein products of this gene are useful for study, treatment, and diagnosis of many cancer types, including testicular cancer; as well as disorders involving endocrine function and normal growth and development.

20

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The translation product of this gene shares sequence homology with BID-apoptotic death gene (mouse), Genbank accession no. PID g1669514, which is thought to be important in programmed cell death.

25

This gene is expressed primarily in jurkat membrane bound polysomes and activated neutrophils and to a lesser extent in endothelial cells and human cerebellum.

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers and other proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and 35 cell types (e.g., blood cells, endothelium, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

- urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
5 NO:155 as residues: Glu-4 to Leu-11, Cys-28 to Arg-35, Gln-50 to His-66, Glu-73 to Gln-79, Gly-94 to Ser-100, Arg-114 to Asp-126, Pro-139 to Lys-146.
- The tissue distribution and homology to BID-apoptotic death gene indicates that the protein products of this gene are useful for study of cell death, and treatment and diagnosis of proliferative disorders and cancers. Apoptosis - programmed cell death - is  
10 a physiological mechanism involved in the deletion of peripheral T lymphocytes of the immune system, and its dysregulation can lead to a number of different pathogenic processes. Diseases associated with increased cell survival, or the inhibition of apoptosis, include cancers (such as follicular lymphomas, carcinomas with p53 mutations, and hormone-dependent tumors, such as breast cancer, prostate cancer,  
15 Kaposi's sarcoma and ovarian cancer); autoimmune disorders (such as systemic lupus erythematosus and immune-related glomerulonephritis rheumatoid arthritis) and viral infections (such as herpes viruses, pox viruses and adenoviruses), inflammation; graft vs. host disease, acute graft rejection, and chronic graft rejection. Diseases associated with increased apoptosis include AIDS; neurodegenerative disorders (such as  
20 Alzheimer's disease, Parkinson's disease, Amyotrophic lateral sclerosis, Retinitis pigmentosa, Cerebellar degeneration); myelodysplastic syndromes (such as aplastic anemia), ischemic injury (such as that caused by myocardial infarction, stroke and reperfusion injury), toxin-induced liver disease (such as that caused by alcohol), septic shock, cachexia and anorexia. Thus, the invention provides a method of enhancing  
25 apoptosis in an individual by treating the individual with a polypeptide encoded by this gene.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 32

- 30 The translation product of this gene shares sequence homology with human fructose transporter which is thought to be important in normal metabolic function and activity.
- This gene is expressed primarily in T-cell lymphoma.  
Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, leukemia and other cancers, and metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic, 5 lymph and metabolic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., 10 the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:156 as residues: Pro-22 to Gly-48, Ser-54 to Pro-61.

The tissue distribution indicates that the protein products of this gene are useful for study of mechanisms leading to cancer, treatment and diagnosis of cancerous and 15 pre-cancerous conditions; as well as the study and treatment of various metabolic diseases and disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human meningima. 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and other disorders of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological 25 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., meningima and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell 30 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:157 as residues: Asn-23 to Pro-31.

The tissue distribution indicates that the protein products of this gene are useful 35 for study, diagnosis and treatment of disorders of the CNS and inflammatory responses.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 34**

This gene is expressed primarily in activated monocytes and wound healing tissues and to a lesser extent in fetal epithelium.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and inflammatory disorders and wound healing and tissue repair dysfunctions. Similarly, polypeptides and antibodies directed to these polypeptides are  
10 useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, epithelial and gastrointestinal systems, and healing wounds, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., monocytes and other blood cells, and epithelium, and  
15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:158 as residues:  
20 Ala-28 to Ala-33, Gly-35 to Glu-45.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis, study and treatment of immune and inflammatory disorders and wound healing dysfunctions.

**25 FEATURES OF PROTEIN ENCODED BY GENE NO: 35**

This gene is expressed primarily in human osteosarcoma and prostate cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skeletal and neoplastic conditions such as bone and prostate cancer.  
30 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skeletal systems, expression of this gene at significantly higher or lower  
35 levels may be routinely detected in certain tissues (e.g., bone, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:159 as residues:

5 Ser-14 to Gly-22, Leu-37 to Gln-43.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of skeletal disorders and cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 36

10 This gene encodes a protein which is highly homologous to a protein called congenital heart disease protein 5, presumably implicated in congenital heart disease (see Genbank PID g2810996).

15 This gene is expressed primarily in Hodgkin's lymphoma, erythroleukemia cells, and TNF activated synovial fibroblasts, to a lesser extent in ovarian cancer, cerebellum, spleen, fetal liver and placenta and finally to a lesser extent in various other mesenchymal tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, immune, hematopoietic and cardiovascular disorders. Similarly, 25 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hematopoietic and cardiovascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., heart and other cardiovascular tissue, lymphoid tissue, blood cells, bone marrow, ovary and other reproductive tissue, brain and other tissue of the nervous system, spleen, liver, and mesenchymal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell 30 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:160 as residues: Lys-41 to Met-49, Gln-54 to Glu-59, Glu-76 to Thr-88.

35 The homology of this gene and translation product to congenital heart disease protein 5 indicates a role for this protein in the diagnosis, prognosis and/or treatment of

heart disease or other cardiovascular related disorders. In addition, predominant expression in cells associated with the immune and hematopoietic system indicates a role for this protein in the treatment, diagnosis and/or prognosis of immune and autoimmune diseases, such as lupus, transplant rejection, allergic reactions, arthritis, asthma, 5 immunodeficiency diseases, leukemia, AIDS, thymus disorders such as Graves Disease, lymphocytic thyroiditis, hyperthyroidism and hypothyroidism, graft versus host reaction, graft versus host disease, transplant rejection, myelogenous leukemia, bone marrow fibrosis, and myeloproliferative disease. The protein could also be used to enhance or protect proliferation, differentiation and functional activation of 10 hematopoietic progenitor cells such as bone marrow cells, which could be useful for cancer patients undergoing chemotherapy or patients undergoing bone marrow transplantation. The protein may also be useful to increase the proliferation of peripheral blood leukocytes, which could be useful in the combat of a range of hematopoietic disorders including immunodeficiency diseases, leukemia, and septicemia.

15

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, urogenital neoplasias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at 25 significantly higher or lower levels may be routinely detected in certain tissues (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an 30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:161 as residues: Asn-22 to Asn-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of ovarian and other tumors.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 38**

The translation product of this gene shares sequence homology with zinc finger proteins.

This gene is expressed primarily in various fetal, cancer, and endothelial lines.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for
- 10 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
- 15 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of immune and developmental conditions and cancer.

20

**FEATURES OF PROTEIN ENCODED BY GENE NO: 39**

This gene is expressed primarily in fetal, infant, and adult brain and to a lesser extent in other brain and endocrine organs and blastomas.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain tumors and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of
- 30 disorders of the above tissues or cells, particularly of the nervous and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, endocrine tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
- 35 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

disorder.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of brain cancer and other neurological disorders.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 40

The translation product of this gene shares sequence homology with vesicular glycoproteins and lectins. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: DTYPNEEKQQQERVFPXXSAMVNNGSLSYDHER DGRPTELGGCXAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKHEWRDCIEVPGV  
10 RLPRGYYFGTSSITGDLSDNHDVISLKLFEVTPEEE (SEQ ID NO:281); and/or LKREHSLSKPYQGVGTGSSLWNLMGNAMVMTQYIRLTPDMQSKQGA LWNRVPCFLRDWELQVHFKIHGQGKKNLHGDGLAIWYT (SEQ ID NO:282).

This gene is expressed primarily in infant brain and to a lesser extent in various normal and transformed neural, endocrine, and immune organs.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and neurodevelopmental conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hormonal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, endocrine tissue, and tissue and cells of the immune system, and cancerous and wounded tissues)  
20 or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:164 as residues: Pro-64 to Gly-71, Gly-  
25 94 to Leu-100, Thr-110 to Pro-116, Thr-135 to Arg-145, Glu-164 to Glu-171, Asp-204 to Asp-211, Arg-253 to His-261, Asn-312 to Tyr-323.  
30

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of mental retardation and other neurological disorders and neoplasias.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 41**

This gene displays homology to the glycosyltransferase family, which catalyze the addition of sialic acids to carbohydrate groups which are present on glycoproteins.

5 This gene is expressed primarily in smooth muscle and to a lesser extent in pineal gland, fetal liver, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, gastrointestinal injury, inflammatory and neurodegenerative conditions.

10 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, pineal gland,

15 liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

20 comprising a sequence shown in SEQ ID NO:165 as residues: Ser-12 to Thr-21, Arg-24 to Pro-32, Asp-73 to Lys-82, Lys-90 to Ala-97.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of neurodegenerative and growth disorders and gastrointestinal repair.

25

**FEATURES OF PROTEIN ENCODED BY GENE NO: 42**

The translation product of this gene shares sequence similarity with metallothionein polypeptides. See, for example, Proc. Natl. Acad. Sci. U S A 1992 Jul 15:89(14):6333-6337. Metallothioneins are believed to inhibit neuronal survival among other biological functions. Based on the sequence similarity (especially the conserved cysteine motifs characteristic of the metallothionein family) the translation product of this gene is expected to share certain biological activities with other members of the metallothionein polypeptide family. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: PGTLQCSALHDPGCANCSRFCRD

30 CSPPACQC (SEQ ID NO:283).

35 This gene is expressed exclusively in placenta and fetal liver.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, liver, brain and other tissue of the nervous system, 5 and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 10 15 The tissue distribution and homology to metallothionein indicates that the protein products of this gene are useful for diagnosis and treatment of immune and hematopoietic system disorders and neurological diseases, especially in fetal development.
- 20 **FEATURES OF PROTEIN ENCODED BY GENE NO: 43**
- Preferred polypeptides encoded by this gene comprise the following amino acid sequence: FLYDVLMXHEAVMRTHQIQLPDPEFPS (SEQ ID NO:284).
- This gene is expressed primarily in T-cells and synovial tissue.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of 25 the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, 30 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 35

The tissue distribution indicates that the protein products of this gene are useful for treatment and diagnosis of disorders of the immune system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 44

- 5       The translation product of this gene shares sequence similarity with several methyltransferases (e.g., see Genbank gil1065505).  
This gene is expressed primarily in ovary, thymus, infant adrenal gland, tissues of the nervous system and the hematopoietic tissue, and to a lesser extent in adipose tissue and many other tissues.
- 10      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the reproductive system, the endocrine system, the hematopoietic system and the CNS. Similarly, polypeptides and antibodies directed to 15     these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, endocrine, CNS and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., ovary and other reproductive tissue, thymus, adrenal gland, 20     brain and other tissue of the nervous system, hematopoietic tissue, and adipose tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the 25     disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:168 as residues: Ser-3 to Gly-12, Asp-19 to Arg-31, Tyr-70 to Tyr-77, Asn-130 to Lys-140, Pro-165 to Gln-170, Pro-192 to Lys-199, Leu-216 to Glu-227, Glu-254 to Phe-281.
- 30      The tissue distribution and homology to methyltransferase indicates that the protein products of this gene are useful for diagnosis and treatment of disorders of the CNS, the hematopoietic system and reproductive organs and tissues. For example, the abundant expression in the ovary may indicate that the gene product can be used as a hormone with either systemic or reproductive functions; as growth factors for germ cell maintenance and in vitro culture; as a fertility control agent; remedy for sexual 35     dysfunction or sex development disorders; diagnostics/treatment for ovarian tumors, such as serous adenocarcinoma, dysgerminoma, embryonal carcinoma,

choriocarcinoma, teratoma, etc; The expression in thymus may indicate its utilities in T-cell development and thus its applications in immune related medical conditions, such as infection, allergy, immune deficiency, tissue/organ transplantation, etc.

##### 5    FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with cytochrome C oxidase which is thought to be important in metabolic function of cells. This gene has now recently been published as estrogen response gene. See Genbank accession no. AB007618 and Mol. Cell. Biol. 18 (1), 442-449 (1998). See also J Immunol. Mar 10 1:154(5): 2384-2392 (1995), where the mouse homologue was published and implicated in siliocis.

This gene is expressed primarily in adipose tissue, kidney and fetal brain and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic diseases involving especially adipose tissue, brain and kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adipose tissue, kidney, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:169 as residues: Thr-5 to Ser-14.

The tissue distribution and homology to cytochrome C oxidase, estrogen response gene product and siliocis related gene product indicates that the protein products of this gene are useful for diagnosis and treatment of metabolic disorders in the CNS, adipose tissue and kidney, particularly siliocis.

##### FEATURES OF PROTEIN ENCODED BY GENE NO: 46

35    The translation product of this gene shares sequence homology with reticulocalbin. See, for example, J. Biochem. 117 (5), 1113-1119 (1995). Based on the

sequence similarity, the translation product of this gene is expected to share certain biological activities with reticulocalbin, e.g., Ca++ binding activities. This gene product is sometimes hereinafter referred to as "Reticulocalbin-2".

5 This gene is expressed primarily in breast, endothelial cells, synovial, heart and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the breast, vascular and skeletal/cardiac muscular system.

10 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast, vascular and skeleto-muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g.,  
15 mammary tissue, endothelial cells, synovial tissue, heart and other cardiovascular tissue, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
20 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:170 as residues: Gly-16 to Arg-32, Ala-42 to Asn-50, Glu-66 to Gln-76, Arg-85 to Gly-94, Thr-108 to Asp-115, Trp-121 to Gly-130, Leu-137 to His-144, Glu-155 to Lys-161, Asp-175 to Ser-180, Glu-209 to Gly-217, Glu-232 to Glu-237, Thr-243 to Asp-261, Glu-287 to Arg-295.

25 The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of diseases of the vascular and skeletal/cardiac muscular system. The homology of the gene with reticulocalbin indicates its biological function in regulating calcium store, a particularly important function in muscular cell types. The gene expression in the heart may indicate its utilities in diagnosis and remedy in heart failure, ischemic heart diseases, cardiomyopathy, hypertension, arrhythmia, etc. The abundant expression in the breast may indicate its applications in breast neoplasia and breast cancers, such as fibroadenoma, papillary carcinoma, ductal carcinoma, Paget's disease, medullary carcinoma, mucinous carcinoma, tubular carcinoma, secretory carcinoma and apocrine carcinoma; juvenile hypertrophy and gynecomastia, mastitis  
30 and abscess, duct ectasia, fat necrosis and fibrocystic diseases, etc.  
35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 47**

The translation product of this gene shares weak sequence homology with H<sup>+</sup>-transporting ATP synthase which is thought to be important in cell metabolism or signal transduction.

5 This gene is expressed only in testis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of some types of diseases and conditions.

10 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 Since only one out of about a million expressed sequence tag is found in testes indicates that its expression is low and selectively in testes. Since some of the genes only expressed in testes are usually expressed in brain or in certain induced hematopoietic cells/tissues, it is speculated that this gene to be expressed in brain or hematopoietic cells/tissues and is useful for diagnosis and treatment of disorders these systems.

25

**FEATURES OF PROTEIN ENCODED BY GENE NO: 48**

The translation product of this gene shares sequence homology with human polymeric immunoglobulin receptor (accession No.X73079) which is thought to be important in antibody recognition and immune defenses. In one embodiment, 30 polypeptides of the invention comprise the sequence GWYWCG (SEQ ID NO:285). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta and to a lesser extent in corpus callosum and fetal liver and spleen.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, disorders of the immune system, e.g. autoimmune diseases and immunodeficiency. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:172 as residues: Tyr-37 to Cys-49, Gly-51 to Tyr-56, Lys-88 to Trp-93, Leu-130 to Glu-136.

The tissue distribution and homology to human polymeric immunoglobulin receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, e.g. autoimmune diseases and immunodeficiencies.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 49**

This gene is expressed in thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., thymus and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, e.g. autoimmunity and immunodeficiency.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 50**

Preferred polypeptide encoded by this gene comprise the following amino acid sequence: MKVGARIRVKMSVNKAHPVVSTHWRWPAEWPMFLHLAQEPRTE  
5 VKSRPLGLAGFIRQDSKTRKPLEQETIMSAADTALWPYGHGNREHQENELQKY  
LQYKDMHLLDSGQSLGHTHTLQGSHNLTALNI (SEQ ID NO:286).

Polynucleotides encoding this polypeptide are also provided as are complementary polynucleotides thereto.

10 This gene is expressed primarily in adrenal gland, pituitary, T helper cells, and breast cells and to a lesser extent in a wide variety of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of the some diseases and conditions. Similarly, 15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adrenal gland, pituitary, T-cells and other blood cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids 20 (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:174 as residues: Gln-39 to Ser-47, Arg-57 to Glu-67, 25 Tyr-82 to Gln-95.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of a wide range of disorders, such as immune and endocrine disorders.

**30 FEATURES OF PROTEIN ENCODED BY GENE NO: 51**

The translation product of this gene shares sequence homology with human Sop2p-like protein which is important in cytoskeleton structure. In one embodiment, 35 polypeptides of the invention comprise the sequence SLHKNSVSQISVLSGGKAKCS QFCTTGMDGGMSIWDVKSLESALKDLKI (SEQ ID NO:287). Polynucleotides encoding this polypeptide are also encompassed by the invention. This gene maps to chromosome 7. Therefore, polynucleotides of the invention can be used in linkage

analysis as a marker for chromosome 7.

This gene is expressed primarily in immune and hematopoietic tissues/cells and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and hematopoietic disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a  
10 number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune and hematopoietic tissue/cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:175 as residues: Lys-49 to Gln-54, Ala-61 to Arg-66, Lys-82 to Lys-87, Glu-126 to Val-133, His-136 to Ile-141, Glu-175 to Ser-187, Asp-  
20 286 to Leu-296, Ala-298 to Ser-310.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immunological, hematopoietic, and inflammatory disorders, e.g. immunodeficiency, autoimmunity, inflammation.

25

## FEATURES OF PROTEIN ENCODED BY GENE NO: 52

The translation product of this gene shares sequence homology with *Caenorhabditis elegans* R53.5 gene encoding a putative secreted protein without known function.

- 30 This gene is expressed primarily in endothelial cells, brain and several highly vascularized, and tumor tissues and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, aberrant angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular and brain system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, brain and other tissue of the nervous system, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:176 as residues: Thr-43 to Asn-60, Thr-106 to Phe-115, Asp-122 to Arg-133, Arg-186 to Asp-192, Leu-211 to Lys-216.

The tissue distribution and homology to a *C. elegans* secreted protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of disorders in vascular or brain system, e.g. aberrant angiogenesis, ischemia, neurodegeneration, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 53

In one embodiment, polypeptides of the invention comprise the sequence EASKSSHAGLDLFSVAACHRF (SEQ ID NO:288). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in T-cells and to a lesser extent in brain. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphocytic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphoid system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:177 as residues: Pro-3 to Thr-8, Arg-37 to Asp-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment, and cure of lymphocytic disorders.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with secreted cartilage matrix protein, a major component of the extracellular matrix of nonarticular cartilage which is thought to be important in cartilage structure. In specific embodiments, polypeptides of the invention comprise the sequence: RCKKCTEGPI  
10 DLVFVIDGSKSLGEENFEVVKQF (SEQ ID NO:297); VTGIIIDSLTISPKAARVGL  
LQYSTQVH (SEQ ID NO:290); TEFTLRNFNSAKDMKKAVAHMKYM (SEQ ID  
NO:291); GKGSMTGLALKHMFERSFTQGEGARPF (SEQ ID NO:292); STRVP  
RAAIVFTDGRAQDDVSEWASKAKANGITMYAVGVGKAIE (SEQ ID NO:293);  
EELQEIASEPTNKHLFYAEDFSTMDEISEKLKGICEALEDS (SEQ ID NO:294);  
15 TQRLEEMTQRM (SEQ ID NO:295); PQGCPEQPLH (SEQ ID NO:296); and/or  
YMGKGSMTGLALKHMFERSFT (SEQ ID NO:289). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta, infant brain, prostate, fetal lung and to a lesser extent in endometrium and fetal tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormal placenta and pregnancy, disorder and injury in brain, prostate, and vasculature. Similarly, polypeptides and antibodies directed to these polypeptides  
25 are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction, neuronal, and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, brain and other tissue of the nervous system, prostate, lung and  
30 endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution and homology to cartilage matrix protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis,

treatment, and cure of abnormalities in placenta and pregnancy, disorder and injury in brain, prostate, and vasculature.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 55**

5       The translation product of this gene is the human ortholog of bovine and hamster CII-3, a succinate-ubiquinone oxidoreductase complex II membrane-intrinsic subunit, which is thought to be important in mitochondrial electron transport chain during metabolism. In specific embodiments, the polypeptides of the invention comprise MAALLLRHVGRHCLRAHFSPQLCIRNAVPLGTTAKEEMERFWNKNIG  
10 SNRPLSPHITIYS (SEQ ID NO:298); VFPLMYHTWNGIRHLMWDLGKGLKIPQL YQSG (SEQ ID NO:299); MAALLLRHVGRHCLRAH (SEQ ID NO:300); VKSLCL GPALIHTAKFAL (SEQ ID NO:301); VFPLMYHTWNGIRHLMWDLGKGL (SEQ ID NO:302).

This gene is expressed in 8-week old early stage human.

15      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolism disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
20 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or  
25 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment, and cure of metabolism disorders.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 56**

This gene is expressed primarily in umbilical vein endothelial cells, human ovarian tumor cells, human meningoma cells, and human Jurkat membrane bound  
35 polysomes. In specific embodiments, polypeptides of the invention comprise the amino acid sequence: RVWDVRPFAPKERCVKIFQGNV (SEQ ID NO:303); HNFEKNLL

RCSWSPDGSKIAAGSADRFVYV (SEQ ID NO:304); and/or WDTTSRRILYKLPG HAGSINEAFHPDEPI (SEQ ID NO:305). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, immune and cardiovascular disorders and urogenital neoplasias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neurological, urogenital, reproductive system and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, cells, endothelial cells, ovary and other reproductive tissue, meningima, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:143 as residues: Phe-71 to Arg-76, Pro-82 to His-87, Glu-103 to Ala-111.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immune, immuno-suppressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. In addition, expression in ovarian tumor cells suggests that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of ovarian tumors, and other tumors and neoplasias. Further, endothelial cell expression suggests a role in cardiovascular or respiratory/pulmonary disorders or infections (athsma, pulmonary edema, pneumonia).

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 57

The translation product of this gene shares sequence homology with type I collagen. In specific embodiments, the polypeptides of the invention comprise the sequence: GRIPAPAPSVPAGPDSR (SEQ ID NO:309); VRGRTVLRPGLDAEPE LSPE (SEQ ID NO:306); EQRVLERKLKKERKKEERQ (SEQ ID NO:307); ARRSG

AELAWDYLRCRWAQKHKNWRFQKTRQTWLLLHMYDSDKVPDEHFSTLLAYLE  
GLQGR (SEQ ID NO:255); and/or RLREAGLVAQHPP (SEQ ID NO:308).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in epididymus, prostate cell line (LNCAP),  
5 and pituitary gland; and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the epididymus, prostate (especially prostate cancer),  
10 and pituitary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system and neuroendocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain  
15 tissues (e.g., epididymus and other reproductive tissue, prostate, and pituitary gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the  
20 disorder.

The tissue distribution and homology to type I collagen, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the epididymus, prostate (especially prostate cancer), and pituitary gland.  
25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 58

This gene is expressed primarily in the frontal cortex of the brain from a schizophrenic individual.

Therefore, polynucleotides and polypeptides of the invention are useful as  
30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
35 particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of

the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of schizophrenia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 59

10 The polypeptide encoded by Gene 59 is homologous to human surface 4 integral membrane protein. In specific embodiments, the polypeptides of the invention comprise the sequence: TGCVLVLSRNFVQYACFGLFGIIALQTIAYSILWDLKF LMRN (SEQ ID NO:310); SRSEGKSMFAGVPTMRESSPKQYMQLGGRVLLV LMFMTLLHFDASFFSIVQNIVG (SEQ IDNO:311); GTAEDFADQFLRVTKQYLP  
15 HVARLCLISTFLEDGIRMFQWSEQRDYIDTTWNCGYLLAS (SEQ ID NO:312); LMRNESRS (SEQ ID NO:314); ASFLLSRTSWGTA (SEQ ID NO:315); and/or ASFLLSRTSWGTALMIL (SEQ ID NO:313). Polynucleotides encoding these polypeptides are also encompassed by the invention.

20 This gene is expressed primarily in Hodgkin's lymphoma and lung; and to a lesser extent in many other human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma, tumors or other abnormalities of the lung.

25 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., lymphoid tissue, and  
30 pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
35 NO:183 as residues: Met-20 to Trp-27.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of Hodgkin's lymphoma, tumors or other abnormalities of the lung.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 60**

5 This gene is expressed primarily in bone cancer and stomach cancer, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are  
10 not limited to, bone cancer and stomach cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, and the stomach, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues  
15 (e.g., bone, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone cancer and stomach cancer and possibly other cancers.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 61**

25 This gene is expressed primarily in epididymus, and lymph node of breast cancer, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are  
30 not limited to, abnormalities of the epididymus, and breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epididymus and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., epididymus and other reproductive tissue, lymphoid tissue, and mammary tissue, and cancerous and wounded tissues) or bodily fluids  
35

(e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:185 as residues: Arg-57 to Ser-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the epididymus, and breast cancer.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 62

The translation product of this gene appears to be the human homolog of bovine NADH dehydrogenase which is thought to be important in cellular metabolism. In specific embodiments, the polypeptides of the invention comprise the amino acid sequence: SMSALTRLASFARVGGRFLRSGCARTAGDGGVRHAGGGVHIEPRY RQFPQLTRSQVFQSEFFSGLMWFWILWRFWHDSEEVLGHFPYPDPSQWTDEEL GIPPDED (SEQ ID NO:323), or fragments thereof. Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed in larynx tumor, lymph node, brain amygdala, human cardiomyopathy, and retina.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases affecting cellular metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes 25 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., larynx, lymphoid tissue, brain and other tissue of the nervous system, heart and cardiovascular tissue, and retina, and cancerous and wounded tissues) or 30 bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Pro-27 to Gln-32, Arg- 35 42 to Glu-51.

The tissue distribution and homology to NADH dehydrogenase indicates that

polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases involving cellular metabolism.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 63**

- 5 This gene is expressed primarily in amygdala, and to a lesser extent in many other tissues.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the amygdala. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., 10 amygdala, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a 15 sequence shown in SEQ ID NO:187 as residues: Gln-17 to Glu-29, Pro-41 to Phe-46, Ser-59 to Ile-70, Thr-97 to Leu-105.
- 20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of amygdala.

25

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

This gene is expressed primarily in female bladder, and to a lesser extent in chronic synovitis and hemangiopericytoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as 30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bladder cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary tract, expression of this gene at significantly higher or lower 35 levels may be routinely detected in certain tissues (e.g., bladder, synovial tissue, and

vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:188 as residues: Pro-2 to Gln-7, Pro-27 to Phe-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatments of defects of the urinary tract, especially bladder cancer.

10

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 65**

This gene is expressed primarily in fetal spleen, and to a lesser extent in hemangiopericytoma, thymus, and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, defects of immune or hematopoietic systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., spleen, vascular tissue, thymus, blood cells, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The protein product of this gene is useful for treatment of defects of the immune or hematopoietic systems, because of the gene's expression in thymus and spleen.

30

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 66**

This gene is expressed primarily in human pituitary and to a lesser extent in placenta and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, endocrine growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pituitary and other endocrine tissue, placenta, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:190 as residues: Val-38 to Asn-44, Gly-53 to Ser-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of growth disorders related to pituitary dysfunction.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 67

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene of unknown function. In specific embodiments, the polypeptides of the invention comprise the sequence: DPRRPNKVLRYKPPPSE CNPALDDPTP (SEQ ID NO:317); DYMNLLGMIFSMCGLMLKLKWCAWVA VYCS (SEQ ID NO:318); FISFANSRSSEDTKQMMSSF (SEQ ID NO:316); and/or MLSISAVVMSYLNQPQPMTPPW (SEQ ID NO:319). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in primary breast cancer and lymph node breast cancer and to a lesser extent in adult brain, lung cancer, colon cancer, epithelioid sarcoma, and Caco-2 cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, lymphoid tissue, brain and other tissue of the nervous system, lung, colon, and

epithelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:191 as residues: Asn-34 to Lys-42.

The tissue distribution in a variety of cancer tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of a variety of cancer and tumor types.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 68

The translation product of this gene shares sequence homology with steroid membrane binding protein. The translation product of this gene has recently been published as progesterone binding protein. See Genbank AJ002030. Preferred 15 polypeptides encoded by this gene comprise the following amino acid sequence:  
AAGDGDVKLGLGSGSESSNDGGSESPGDAGAAAXGGGWAAAALALLTG  
GGE (SEQ ID NO:320).

This gene is expressed primarily in breast, and to a lesser extent in placenta and fetal tissue.

20

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer or developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes 25 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of breast or fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, placenta, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another 30 tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:192 as residues: Pro-43 to Asp-49, Gln-54 to Pro-64, Asp-110 to Asp-118, Lys-138 to Tyr-143, Pro-150 to Asp-170.

35

The tissue distribution and homology to steroid membrane binding protein and to progesterone binding protein indicates that the protein products of this gene are

useful for treatment of breast cancers, especially those caused by estrogen and progesterone binding.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 69

- 5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence: AADNYGIPRACRNSARSYGAAWLLLXPAGSSRVEPTQDISISDQLGG QDVPVFRNLSLLVVGVGAVFSLFHLGTRERRPHAXEPGEHTPLAPATAQPL LLWKHWLREXAFYQVGILYMTTRLIVNLSQTYMAMYLTYSLHLPKKFIATIPLV MYLSGFLSSFLMKPINKCIGRN (SEQ ID NO:321).
- 10 This gene is expressed primarily in macrophage (GM-CSF treated), and to a lesser extent in monocytes and dendritic cells.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
- 15 not limited to, inflammation and infection . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell
- 20 types (e.g., macrophages and other blood cells, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 25 The tissue distribution indicates that the protein products of this gene are useful for treatment of infection or inflammation or other events or defects involving the immune system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 70

- 30 This gene is expressed primarily in adult brain and to a lesser extent in thyroid, 12 week old early stage human, and stromal cell TF274.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
- 35 not limited to, neurological or neuro-endocrine diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, 5 thyroid, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a 10 sequence shown in SEQ ID NO:194 as residues: Pro-65 to Cys-71.

The tissue distribution indicates that the protein products of this gene are useful for treatment and diagnosis of neurological diseases or metabolic conditions involving the neuro-endocrine system.

## 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed in T-cell helper and to a lesser extent in adult brain and adult testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, meningitis or reproductive problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neural 25 and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, brain and other tissue of the nervous system, testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual 30 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:195 as residues: Val-18 to Tyr-24, Ala-89 to Asp-99, Asp-104 to Ala-117, Leu-121 to Pro-136.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis immune and

reproductive disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 72

The translated polypeptide of this contig has a high degree of identity with the  
5 Ob Receptor-Associated Protein deposited as GenBank Accession No. 2266638. No  
function has been determined for the Ob Receptor-Associated Protein, however it is  
expressed upon stimulation of the Ob Receptor by Leptin.

This gene is expressed in T-cells and to a lesser extent in endothelial and bone  
marrow cells.

10 Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, acute lymphoblastic leukemia, hematopoietic disorders. Similarly,  
polypeptides and antibodies directed to these polypeptides are useful in providing  
15 immunological probes for differential identification of the tissue(s) or cell type(s). For a  
number of disorders of the above tissues or cells, particularly of the immune and  
hematopoietic systems, expression of this gene at significantly higher or lower levels  
may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood  
cells, endothelial cells, and bone marrow, and cancerous and wounded tissues) or  
20 bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another  
tissue or cell sample taken from an individual having such a disorder, relative to the  
standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder. Preferred epitopes include those  
comprising a sequence shown in SEQ ID NO:196 as residues: Ser-61 to Trp-70.

25 The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for treatment and diagnosis of leukemia and other  
disorders of the primary immune system. In addition, since this gene appears to be  
related to the Ob Receptor-Related Protein, it is likely that this polypeptide is also  
involved in the Ob/Leptin signal transduction cascade. As a result, this protein may be  
30 of use in the molecular diagnosis and therapeutic intervention of obesity and related  
disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 73

The translation product of this contig has homology with furin, a protein  
35 thought to be a key endopeptidase in the constitutive secretory pathway. The  
identification and initial characterization of Furin was reported by Takahasi and

colleagues (Biochem Biophys Res Commun 1993 Sep 15;195(2):1019-1026).

This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system such as allergies, wound healing and antigen recognition. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of allergies or other immune disorders since neutrophils are an important part of an allergic response. Further, since this protein appears to be related to Furin, it can be used diagnostically and therapeutically to treat secretory protein processing disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed in the frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, of the motor activity and sensory functions that involve the central nervous system . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neural disorders  
5 that affect cognitive functions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 75

The translation product of this gene shares sequence homology with inorganic pyrophosphatase which is thought to be important in the catalysis the hydrolysis of

10 diphosphate bonds, chiefly in nucleoside di- and triphosphates and essential enzymes that are important for controlling the cellular levels of inorganic pyrophosphate (PPi). The bovine homolog of this gene has been identified by Yang and Wensel (J. Biol. Chem. 267:24641-24647 (1992)).

15 This gene is expressed in osteoclastoma cells and to a lesser extent in epithelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoporosis and other bone weakening diseases. Similarly, polypeptides

20 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, and epithelial cells, and cancerous and wounded 25 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:199 as residues: Lys-22 to Tyr-28, Asp-

30 64 to Lys-77, Pro-86 to Ile-91, Gln-99 to Pro-119, Tyr-169 to Asp-174, Lys-176 to Gly-181, Trp-189 to Asn-202, Lys-233 to Gly-239, Ser-250 to Asp-257.

The tissue distribution and homology to inorganic pyrophosphatase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of osteoporosis through the removal of bone by demineralization.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 76**

The translation product of this gene shares exact sequence homology with ATP sulfurylase/APS kinase (GenBank Accession No. 2673862) which is thought to be important in biosynthesis of the activated sulfate donor, adenosine 3'-phosphate 5'-phosphosulfate, involves the sequential action of two enzyme activities: ATP sulfurylase, which catalyzes the formation of adenosine 5'-phosphosulfate (APS) from ATP and free sulfate, and APS kinase, which subsequently phosphorylates APS to produce adenosine 3'-phosphate 5'-phosphosulfate.

This gene is expressed in osteoclastoma cells and to a lesser extent in  
10 developmental tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, antibiotic resistant bacterial infections, osteoarthritis and other auto

15 immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or skeletal structure expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and  
20 developmental tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a  
25 sequence shown in SEQ ID NO:200 as residues: Asn-15 to Trp-20, Ser-36 to Gly-41, Pro-103 to Val-110, Pro-134 to Arg-143, Leu-173 to Arg-178, Ser-190 to Ala-197, His-314 to Arg-319, Arg-354 to Asn-362, Asp-391 to Arg-397, Glu-402 to Asp-409, Asp-434 to Leu-439, Glu-441 to Arg-446, Gly-455 to Asp-462, Pro-528 to His-541, Asn-566 to Arg-571, Tyr-574 to Glu-581, Thr-589 to Glu-603.

30 The tissue distribution and homology to ATP sulfurylase/APS kinase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment or detection of autoimmune diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 77**

35 This polypeptide is identical to the SLP-76-associated protein reported by Musci and colleagues (J. Biol. Chem. 272 (18), 11674-11677 (1997)) and to the FYB protein

reported by da Silva and coworkers (Proc. Natl. Acad. Sci. U.S.A. (1997) In press).

These proteins have been reported to be novel T-cell Proteins which bind FYN and SLP-76 and regulate IL-2 production. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: RITDNPEGKWLGRRTARGSYGYIK

5       TTAVEIXYDSLKLKKDSLAPSRIEDDQEYDDVAEQDDISHSQSGSGGIFPP  
PPDDDIYDGIEEEDADDGFPAPPKQLDMGDEVYDDVTSDFPVSSAEMSQGTNV  
GKAKTEEKDLKKLKKQXKEKDFRKFKYDGEIRVLYSTKVTTSITSKKWGT  
RDLQVKPGESLEVIQTTDDTKVLCRNEEGKYGYVLRSYLADNDGEIYDDIADGC  
IYDND (SEQ ID NO:322).

10       This gene is expressed in CD34 positive cells (hematopoietic progenitor cells) and to a lesser extent in adult spleen derived from a chronic lymphocytic leukemia patient.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, chronic lymphocytic leukemia; hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., T-cells and other blood cells, bone marrow, hematopoietic cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Further, nucleic acids and polypeptides of the present invention are useful both diagnostically and therapeutically in the intervention of immune and other disorders in which the ability to alter IL-2 expression is desired. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:201 as residues: Ala-17 to Lys-37, Val-39 to Ser-45, Lys-59 to His-70, Arg-90 to Leu-95, Lys-97 to Lys-107, Ser-117 to Leu-124, Phe-133 to Ser-138, Trp-146 to Leu-167, Pro-175 to Asn-185, Lys-190 to Ser-211, Pro-213 to Ser-222, His-230 to Pro-235, Pro-240 to Pro-246, Pro-253 to Gly-261, Leu-271 to Leu-303, Leu-305 to Leu-326, Lys-343 to Leu-349, Thr-363 to Leu-371, Arg-373 to Tyr-381, Tyr-391 to Leu-401, Pro-404 to Val-414, Ser-426 to Ser-432, Ile-448 to Ser-457, Gln-462 to Trp-468, Lys-477 to Ser-501, Asp-518 to Ser-523, Ala-541 to Gln-554.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of a variety of hematopoietic disorders. The noted expression of this gene in the hematopoietic progenitor cell compartment - as determined by its expression on CD34 positive hematopoietic stem 5 and progenitor cells - indicates that it plays a critical role in the expansion or proliferation of hematopoietic stem/progenitor cells, as well as in the differentiation of the various blood cell lineages. Thus it could be useful in the reconstitution of the hematopoietic system of patients with leukemias and other hematopoietic diseases.

**10 FEATURES OF PROTEIN ENCODED BY GENE NO: 78**

This gene is homologous to heparin cofactor II (HCII) which is a 66-kDa plasma glycoprotein that inhibits thrombin rapidly in the presence of dermatan sulfate or heparin.

This gene is expressed in apoptotic and anergic T-cells.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thrombopenia T-cell lymphomas; Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing 20 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system - most notably the T-cell compartment, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily 25 fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The homology to heparin cofactor II (HCII) and the tissue distribution indicates 30 that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic disorders particularly in thrombopoiesis, most notably of the T-cell compartment. This could include immune modulation, inflammation, immune surveillance, graft rejection, and autoimmunity.

**35 FEATURES OF PROTEIN ENCODED BY GENE NO: 79**

The translation product of this gene shares sequence homology with a mouse

protein believed to represent an integral membrane protein.

This gene is expressed in fetal cochlea and epididymus and to a lesser extent in adult spleen and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma; disorders of the inner ear; male fertility disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inner ear; male reproductive tract; bone; and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cochlea, epididymus and other reproductive tissue, spleen, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:203 as residues: Lys-13 to Gly-23, Cys-38 to Asp-43, Gly-48 to Trp-53, Cys-223 to Ile-237, Ile-240 to Ser-246.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of hearing and fertility disorders. Likewise, it may have a role in the modulation of immune function and in the treatment of osteoporosis.

25

## FEATURES OF PROTEIN ENCODED BY GENE NO: 80

The translation product of this gene shares sequence homology with reticulocalbin which is thought to be important in the binding of calcium, particularly within the endoplasmic reticulum.

30 This gene is expressed in endothelial cells and stromal cells and to a lesser extent in osteoblasts, osteoclasts, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoperosis; osteoclastomas; T-cell lymphomas; Hodgkin's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

- providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature, bone, and immune systems - particularly the T-cell compartments, expression of this gene at significantly higher or lower levels may be routinely detected
- 5 in certain tissues and cell types (e.g., endothelial cells, stromal cells, bone, T-cells and other blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
- 10 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:204 as residues: Lys-20 to Arg-27, Pro-32 to Asp-48, Leu-64 to Arg-72, Asp-108 to Lys-114, Glu-128 to Thr-133, Asp-139 to Phe-147, Thr-196 to Ala-204, Tyr-218 to Glu-228, Val-230 to Gln-236, Arg-241 to Lys-255, Glu-276 to Lys-287.
- 15 The tissue distribution and homology to reticulocalbin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of bone disorders such as osteoporosis; the diagnosis and treatment of T-cell lymphomas and Hodgkin's lymphoma; and the treatment of diseases and defects of the vasculature, such as vascular leak syndrome and aberrant
- 20 angiogenesis that accompanies tumor growth.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 81

The translation product of this gene shares sequence homology with a family of peptide transport genes - particularly the AtPTR2-B gene from *Arabidopsis* - which are thought to be important in the uptake of small peptides.

This gene is expressed in a number of fetal tissues, most notably lung, brain, cochlea, and liver/spleen, and to a lesser extent in osteoclastoma and endometrial tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma; endometrial tumors; cancer; leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and endometrium, expression of this gene at significantly higher or lower levels may be

- routinely detected in certain tissues (e.g., fetal tissue, pulmonary tissue, bone, brain and other tissue of the nervous system, cochlea, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,
- 5 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:205 as residues: Lys-186 to Asn-199, Pro-202 to Ala-207.
- The tissue distribution and homology to peptide transport genes indicates that
- 10 polynucleotides and polypeptides corresponding to this gene are useful for the control of cell proliferation, owing to its strong expression in fetal tissues undergoing active cell division, as well as its expression in a variety of tumors or cancers of adult tissues. Potentially, it may regulate the uptake of peptides that stimulate cell proliferation. This gene product may also be useful in stimulating the uptake of a variety of peptide-based
- 15 drug compounds.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 82

- This gene is expressed in fetal liver and spleen and to a lesser extent in endothelial cells.
- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and tumors of a hematopoietic and/or endothelial cell origin; leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are
- 25 useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and/or vasculature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, endothelial cells, vascular tissue, and tissue and cells of the immune system,
- 30 and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
- 35 NO:206 as residues: Met-1 to Asp-9, Arg-66 to Gly-76, Asp-164 to Arg-171.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for the treatment of disorders of the immune system. Expression of this gene product in both fetal liver/spleen and endothelial cells indicates that it may be expressed in the hemangioblast, the progenitor cell for both the immune system and the vasculature. Thus, it is most likely expressed in hematopoietic stem cells, and may be useful for the expansion of hematopoietic stem and progenitor cells in conjunction with cancer treatment for a variety of leukemias.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 84

The translation product of this gene shares sequence homology with NADH dehydrogenase which is thought to be important in cellular metabolism.

This gene is expressed in fetal dura mater and to a lesser extent in T-cells and hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases affecting cellular metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, T-cells and other blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Pro-27 to Gln-32, Arg-42 to Glu-51.

The tissue distribution and homology to NADH dehydrogenase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases involving cellular metabolism.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 85**

The translation product of this gene shares sequence homology with I-TRAF, a novel TNF receptor associated factor (TRAF)-interacting protein that regulates TNF receptor-mediated signal transduction. This protein is thought to be important in regulating the cellular response to tumor necrosis factor (TNF), which is an important mediator of inflammation.

This gene is expressed in endothelial cells and to a lesser extent in glioblastoma and osteoblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation; glioblastoma and osteoblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, bone, and glial cells and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:209 as residues: Glu-15 to Thr-22, Glu-46 to Leu-62, Arg-103 to Glu-119, Gln-127 to Glu-132, Asn-152 to Trp-158, Gln-191 to Gln-210, Glu-264 to Thr-271, Tyr-282 to Leu-288, Trp-319 to Thr-331, Glu-335 to Ser-348, Ser-353 to Ser-358, Asp-382 to Asn-392.

The tissue distribution and homology to I-TRAF indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, including rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, particularly where tumor necrosis factor is known to be involved.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 86**

This gene has homology with a candidate gene involved in X-linked Retinopathy reported by Wong and colleagues (Genomics 15:467-471 (1993)).

This gene is expressed in a T-cell line.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and autoimmune diseases; T-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
- 15 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory disorders such as sepsis, inflammatory bowel disease, psoriasis, and rheumatoid arthritis as well as autoimmune disease such as lupus. It could also be useful in immune modulation and in the process of immune surveillance. The present invention can be used diagnostically and therapeutically to treat X-linked Retinopathy.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 87**

This gene is expressed in human brain tissue.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain disorders; neurodegenerative disorders; tumors of a brain origin.
- 30 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels
- 35 may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:211 as residues: Cys-32 to Tyr-38.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of CNS disorders such as epilepsy, paranoia, depression, Alzheimer's disease, and schizophrenia. It could be useful in the survival and/or proliferation of neurons and could effect neuronal regeneration.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HAGEW82	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	11	1679	247	1607	353	353	125	1
2	HAGFY16	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	12	1830	87	1786	128	128	126	1
2	HBMCF37	xxxxx 03/19/98	pBluescript	98	1487	79	1487	170	170	212	1
2	HFLQBI6	209641 02/25/98	Uni-ZAP XR	99	1653	394	1637	413	413	213	1
3	HALAA60	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	13	1212	1	1212	99	99	127	1
4	HAPBL78	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	14	2061	882	2061	900	900	128	1
5	HASA V70	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	15	1412	10	733	103	103	129	1

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq. X	3' NT of Clone Seq.	NT SEQ ID NO: Y	5' NT of AA SEQ ID NO: Y	First AA of Signal Pep	Last AA of Sig Pep	First AA of Secreted Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
6	HBNAF22	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	16	1052	276	880	538	538	130	1	17	18
7	HBNBL77	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	17	683	1	683	181	181	131	1		29
8	HCDDR90	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	18	1054	86	1007	86	86	132	1	23	24
9	HCEEF50	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	19	1393	132	1393	192	192	133	1	17	18
10	HCEMU42	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	20	1215	277	1070	401	401	134	1	18	19
11	HCENE16	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	21	2042	614	2011	793	793	135	1	26	27

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HMSIJ74	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	22	1872	21	1872	69	69	136	1	23
13	HCUBF15	97923 03/07/97 209071 05/22/97	ZAP Express	23	289	1	289	89	89	137	1	29
14	HE2DE47	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	24	3533	2821	3532	808	808	138	1	30
14	HE2DE47	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	100	1145	435	1115	515	515	214	1	22
15	HKMLH01	209179 07/24/97	pBluescript	25	1148	171	907	196	196	139	1	26
15	HE6DG34	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	101	734	25	734	295	295	215	1	36
16	HE9DG49	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	26	717	1	717	70	70	140	1	27

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
16	HE9DG49	97923 03/07/97 209071 05/22/97	Uni-ZAP XR 102	713	17	713	78	216	1	28	29
17	HELBA06	97923 03/07/97 209071 05/22/97	Uni-ZAP XR 27	1099	1	1099	38	141	1	22	23
17	HELBA06	97923 03/07/97 209071 05/22/97	Uni-ZAP XR 103	1080	1	1080	149	217	1	25	26
18	HSLFM29	97923 03/07/97 209071 05/22/97	Uni-ZAP XR 28	941	171	941	128	142	1	42	43
19	HELBW38	97923 03/07/97 209071 05/22/97	Uni-ZAP XR 29	756	62	756	294	143	1	30	31
20	HETHN28	97923 03/07/97 209071 05/22/97	Uni-ZAP XR 30	2100	408	2093	496	144	1		19

77

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total NT Seq.	3' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
21	HFCDK17	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	31	1448	475	1392	567	145	1	29
22	HFEAF41	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	32	456	1	409	21	146	1	29
23	HFKFL13	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	33	1326	1	1322	210	210	147	1
24	HFSBG13	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	34	710	1	710	242	242	148	1
25	HFTBE43	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	35	1188	110	1161	178	178	149	1
26	HFTDJ36	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	36	956	1	938	144	144	150	1
27	HKTAC77	97924 03/07/97	Uni-ZAP XR	37	1603	974	1581	1104	1104	151	1

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion
28	HLHSH36	97924 03/07/97	pBluescript	38	1089	55	1067	209	152	1		7
29	HLHSV96	97924 03/07/97	pBluescript	39	629	1	629	119	119	153	1	32
30	HLQQBQ86	97924 03/07/97	Lambda ZAP II	40	1964	408	1793	581	154	1		25
31	HLTBX31	97924 03/07/97	Uni-ZAP XR	41	1522	13	1123	126	126	155	1	32
32	HLTCJ63	97924 03/07/97	Uni-ZAP XR	42	875	1	875	43	43	156	1	18
33	HMKAH44	97924 03/07/97	pSport1	43	843	1	843	171	171	157	1	18
34	HMQAJ64	97924 03/07/97	Uni-ZAP XR	44	489	3	489	55	55	158	1	19
34	HMQAJ64	97924 03/07/97	Uni-ZAP XR	104	489	6	489	58	58	218	1	22
35	HOABGG65	97924 03/07/97	Uni-ZAP XR	45	534	1	534	17	17	159	1	18
36	HODCL36	97924 03/07/97	Uni-ZAP XR	46	1374	1	1374	15	15	160	1	20
36	HODCL36	97924 03/07/97	Uni-ZAP XR	105	640	58	640	72	72	219	1	20
37	HODCL50	97924 03/07/97	Uni-ZAP XR	47	596	1	596	269	269	161	1	27

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
38	HODCV74	97924 03/07/97	Uni-ZAP XR	48	851	99	822	170	170	162	1	22
39	HODCZ16	97924 03/07/97	Uni-ZAP XR	49	2020	569	2020	638	638	163	1	17
40	HTOEU03	97924 03/07/97	Uni-ZAP XR	50	2432	848	2432	99	99	164	1	18
40	HTOEU03	97924 03/07/97	Uni-ZAP XR	107	2435	849	2435	928	928	221	1	19
41	HPBCJ74	97924 03/07/97	pBluescript SK-	51	2340	1627	2340	150	150	165	1	20
41	HPBCJ74	97924 03/07/97	pBluescript SK-	108	805	92	791	239	239	222	1	20
42	HPMBU33	97924 03/07/97	Uni-ZAP XR	52	601	188	601	432	432	166	1	20
43	HSAUL66	97924 03/07/97	Uni-ZAP XR	53	359	1	337	142	142	167	1	21
44	HSIDQ18	97924 03/07/97	Uni-ZAP XR	109	1166	21	1166	433	433	223	1	22
45	HSJBB37	97924 03/07/97	Uni-ZAP XR	55	1560	63	1148	217	217	169	1	23
46	HSJBQ79	97924 03/07/97	Uni-ZAP XR	56	1507	164	608	57	57	170	1	24
46	HSJBQ79	97924 03/07/97	Uni-ZAP XR	110	586	4	586	35	35	224	1	25

80

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF		
47	HTEGA76	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	57	450	1	450	83	83	171	1	35	36	68
48	HTEJN13	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	58	1147	1	1147	163	163	172	1	15	16	158
48	HTEJN13	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	111	1134	1	1134	155	155	225	1	19	20	70
49	HTHBL86	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	59	777	1	777	115	115	173	1	18	19	122
50	HTSFO71	97958 03/13/97 209072 05/22/97	pBluescript	60	1191	48	598	52	52	174	1	30	31	128
50	HTSFO71	97958 03/13/97 209072 05/22/97	pBluescript	112	1333	594	1333	829	829	226	1			9
51	HAPNO80	209235 09/04/97	Uni-ZAP XR	61	1580	443	1554	114	114	175	1	1	2	371

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT SEQ D	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
51	HAUCC47	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	113	1015	249	708	244	244	227	1	28
52	HBMCL41	97958 03/13/97 209072 05/22/97	pBluescript	62	1117	105	1034	182	182	176	1	28
53	HCFLD84	97958 03/13/97 209072 05/22/97	pSport1	63	361	1	361	97	97	177	1	32
54	HE8EM69	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	64	1668	1	1638	150	150	178	1	20
55	HE8EZ48	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	65	1353	35	1303	231	231	179	1	33
56	HERGFF73	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	66	1011	655	1011	703	703	180	1	38

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total NT Seq. X	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	AA of Signal Pep	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	Last AA of ORF
57	HFEBF41	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	67	1193	267	1090	459	459	181	1	35	36	95
58	HFRBU14	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	68	560	1	560	63	63	182	1	29	30	94
59	HFV GZ79	97958 03/13/97 209072 05/22/97	pBluescript	69	1657	765	1581	839	839	183	1	21	22	26
60	HHGCM76	97958 03/13/97 209072 05/22/97	Lambda ZAP II	70	711	8	711	270	270	184	1			10
61	HHGCO88	97958 03/13/97 209072 05/22/97	Lambda ZAP II	71	935	111	935	272	272	185	1	19	20	64
62	HHGCP52	97958 03/13/97 209072 05/22/97	Lambda ZAP II	72	504	113	484	127	127	186	1	21	22	21

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
63	HHGDB72	97958 03/13/97 209072 05/22/97	Lambda ZAP II	73	620	1	620	96	187	1	18	19
64	HHGDI71	97958 03/13/97 209072 05/22/97	Lambda ZAP II	74	581	156	581	248	188	1	32	33
65	HHSDI45	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	75	1843	537	1786	630	189	1	27	28
66	HHSEB66	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	76	1441	116	800	167	190	1	36	37
67	HIPAZ83	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	114	1076	398	1076	575	228	1	11	12
68	HLDBO49	97958 03/13/97 209072 05/22/97	PCMV Sport 3.0	78	2776	18	1888	187	192	1	14	15

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	Last AA of ORF		
69	HLDHQ19	97958 03/13/97 209072 05/22/97	PCMVSport 3.0	79	1525	401	1480	534	534	193	1	22	23	65
69	HLDHQ19	209226 08/28/97	PCMVSport 3.0	115	1487	401	1487	534	534	229	1	22	23	131
70	HMSGT42	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	80	1563	33	1077	40	40	194	1	32	33	91
71	HMWIC78	97957 03/13/97 209073 05/22/97	Uni-Zap XR	81	1020	18	780	238	238	195	1	23	24	175
72	HTTCT79	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	82	770	101	770	286	286	196	1	26	27	69
73	HNGJU84	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	83	481	1	481	58	58	197	1	20	21	24
74	HNTAC73	97957 03/13/97 209073 05/22/97	PCMVSport 3.0	84	644	1	623	14	14	198	1	25	26	72

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SBQ ID NO: Y	First AA of Signal Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF		
75	HOSEI45	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	85	1351	435	1284	98	98	199	1	12	13	288
75	HOSEI45	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	116	1350	428	1283							27
76	HOSFD58	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	86	2527	290	1747	56	56	200	1	30	31	623
76	HOSFD58	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	117	2527	288	1747	477	477	231	1	32	33	60
77	HSAUM95	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	87	2566	1843	2566	251	251	201	1	30	31	648
77	HSAUM95	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	118	1098	375	1098	677	677	232	1	21	22	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT of Total NT Seq. X	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	Last AA of ORF	
78	HSAUR67	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	88	540	1	540	83	83	202	1	32	33	54
79	HSKDI81	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	89	1863	152	1165	188	188	203	1	11	12	265
79	HSKDI81	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	119	1679	152	1166	315	315	233	1			17
80	HSKDW91	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	90	2478	1149	2449	92	92	204	1	19	20	314
81	HTLEX50	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	91	2058	476	2058	414	414	205	1	20	21	206
82	HSKHL65	97957 03/13/97 209073 05/22/97	pBluescript	92	1411	345	1411	157	157	206	1	69	70	194

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
82	HSKHL65	97957 03/13/97 209073 05/22/97	pBluescript	121	1411	345	1411	526	235	1	37	38
83	HHFGA11	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	93	2187	147	2184	397	397	207	1	30
83	HHFGA11	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	122	2256	138	2063	228	228	236	1	19
84	HWTBL40	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	94	757	524	608	445	445	208	1	20
85	HBXFG80	97957 03/13/97 209073 05/22/97	ZAP Express	95	2394	481	2394	523	523	209	1	1
86	HCACY32	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	96	672	1	672	117	117	210	1	21

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ NO: X	5' NT of Total NT Seq. X	3' NT of Clone Seq.	5' NT of Clone Seq.	AA of Start Codon Pep	AA of Signal Pep	AA of Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	Last AA of ORF
87	HCEDO21	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	97	1419	1	1419	207	207	211	1	20	21	37

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification , such as multiple histidine residues, or an additional sequence for stability during recombinant production.

- 5       The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988).  
Polypeptides of the invention also can be purified from natural or recombinant sources  
10      using antibodies of the invention raised against the secreted protein in methods which  
are well known in the art.

### Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the  
15      cleavage point for that sequence, are available. For instance, the method of McGeoch,  
Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged  
region and a subsequent uncharged region of the complete (uncleaved) protein. The  
method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information  
from the residues surrounding the cleavage site, typically residues -13 to +2, where +1  
20      indicates the amino terminus of the secreted protein. The accuracy of predicting the  
cleavage points of known mammalian secretory proteins for each of these methods is in  
the range of 75-80%. (von Heinje, *supra*.) However, the two methods do not always  
produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide  
25      was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein  
Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on  
the amino acid sequence. As part of this computational prediction of localization, the  
methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid  
sequences of the secreted proteins described herein by this program provided the results  
30      shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes  
vary from organism to organism and cannot be predicted with absolute certainty.  
Accordingly, the present invention provides secreted polypeptides having a sequence  
shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., +  
35      or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in  
some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

10 **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between 30 a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result 35 of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

- amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,
- 5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be

10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and

15 subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window

20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.

25 For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of

30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are

35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

- For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired 5 residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence.
- 10 This time the deletions are internal deletions so there are no residues at the N- or C- termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query 15 sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or 20 activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in 25 the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

30 Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be 35 deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham 5 and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the 10 protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues 15 Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, 20 where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino 25 acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins 30 with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 5 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers 10 as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-15 450, 451-500, 501-550, 551-600, 651-700, or 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

20 In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the 25 invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) 30 amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-35 60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any

combination of the above amino and carboxy terminus deletions are preferred.

Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

- Also preferred are polypeptide and polynucleotide fragments characterized by
- 5 structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.
- 10 Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

30 In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

35 Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et

al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However,

- 5 immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, 10 Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library.

15 Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion 20 proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be 25 used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

30 Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the 35 polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example 5 describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the 10 monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a 15 fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for 20 example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker 25 sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for 30 instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral

- 5 vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is 10 a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to

- 15 name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or 20 UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of

- 25 appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

- 30 Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 35 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods 5 In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, 10 phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also 15 be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production 20 procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the 25 translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

30 **Uses of the Polynucleotides**

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome 35 identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat

polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be

- 5 selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the

- 10 polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome
- 15 specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence *in situ* hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al.,

- 20 "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides

25 correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage

- 30 analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease
- 35 could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the  
5 mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected  
10 individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene  
15 expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science  
20 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model  
25 systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the  
30 present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of  
35 restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying

personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

- 5        The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set  
10      of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as  
15      tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more  
20      restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of  
25      unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

- 30      In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using  
35      DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

- A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.
- In addition to assaying secreted protein levels in a biological sample, proteins can also be detected *in vivo* by imaging. Antibody labels or markers for *in vivo* imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. *In vivo* tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves  
5 (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

- Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S 10 for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).
- 15 Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).
- 20 At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the 25 polypeptides of the present invention can be used to test the following biological activities.

### Biological Activities

- The polynucleotides and polypeptides of the present invention can be used in 30 assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

### Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders  
5 may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in  
10 treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to:  
15 blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also  
20 be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can  
25 decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in  
30 treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation,  
35 differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

- Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, 5 Myasthenia Gravis, Neuropathy, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.
- 10 Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.
- A polynucleotide or polypeptide of the present invention may also be used to 15 treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits 20 an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.
- Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory 25 response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel 30 disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

### Hyperproliferative Disorders

- A polypeptide or polynucleotide can be used to treat or detect hyperproliferative 35 disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary

Syndrome, Waldenström's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the

polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following

DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes

- Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., 5 Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox , hemorrhagic fever, Measles, Mumps,
- 10 15 20 25 30 35 Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.
- Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Nocardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, 5 Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide 10 of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide 15 of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### Regeneration

20 A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal 25 disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and 30 skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase 35 regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue

regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate 5 nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized 10 neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

15 **Chemotaxis**

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of 20 hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. 25 For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present 30 invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

**Binding Activity**

35 A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the  
5 polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology  
1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can  
10 be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed  
15 polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a  
20 labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate  
25 compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The  
30 antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the  
35 polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying 5 agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

### **Other Activities**

10 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

15 A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

20 A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

25 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

30 **Other Preferred Embodiments**

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

35 Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous

- 5 nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete 10 open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

- 15 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

- 20 A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone 25 identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% 30 identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of

- 35 comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

5 Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

10 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

15 Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone 20 identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

25 Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA 30 clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in

5 Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as

10 defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at

15 least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

20 comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence

25 selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an

30 amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is

35 performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide 5 comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10 Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

15 Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid 20 sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human 25 cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an 30 individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of 35 illustration and are not intended as limiting.

ExamplesExample 1: Isolation of a Selected cDNA Clone From the Deposited Sample

5        Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For  
 10      example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
15	Zap Express	pBK
	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSport 2.0	pCMVSport 2.0
	pCMVSport 3.0	pCMVSport 3.0
20	pCR®2.1	pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS.  
 30      The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.  
 35      Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with  $^{32}\text{P}$ - $\gamma$ -ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction  
5 is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl<sub>2</sub>, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are  
10 performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding  
15 portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids  
20 Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the  
25 desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged  
30 RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

35 This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

5

**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

10

**Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

15

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

**Example 4: Chromosomal Mapping of the Polynucleotides**

20

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

35

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

5    **Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as 10 BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>R</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites. 15

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses 20 the lacI repressor and also confers kanamycin resistance (Kan<sup>R</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). 25 The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D. <sup>600</sup>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by 30 centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from 35 QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed 5 with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The 10 recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After 15 renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

25 DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or 30 Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

**Example 6: Purification of a Polypeptide from an Inclusion Body**

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

- 5       Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.
- 10      15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

- 15      The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

- 20      The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

- 25      Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

- 30      To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated  
5 with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for  
10 instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically  
15 the LPS content is less than 0.1 ng/ml according to LAL assays.

**Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System**

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide  
20 into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the  
25 beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

30 Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-  
35 39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a 5 second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

10 The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

15 The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

20 The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

25 Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are 30 added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added.

35 Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life

Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)

- 5 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested
- 10 and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of <sup>35</sup>S-methionine and 5 µCi <sup>35</sup>S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell.

- 25 A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved
- 30 with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),

- 35 pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used

include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the 5 polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing 10 cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); 15 Bebbington et al., Bio/Technology 10:169-175 (1992)). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the 20 expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., 25 with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate 30 restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the 35 naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

5 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for  
10 transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are  
15 trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of  
20 methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

25

#### Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion

proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using  
5 primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can  
be ligated into the BamHI cloning site. Note that the 3' BamHI site should be  
10 destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

15 If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

20 Human IgG Fc region:

GGGATCCGGAGCCAAATCTTCTGACAAAACACACATGCCACC GTGCC  
CAGCACCTGAATTGAGGGTGACCGTCAGTCTCCTCTCCCCCCC AAAACC  
CAAGGACACCTCATGATCTCCGGACTCCTGAGGTACATGCGTGGTGG  
GGACGTAAGCCACGAAGACCTGAGGTCAAGTTCAACTGGTACGTGGACG  
25 GCGTGGAGGTGCATAATGCCAAGACAAAGCCGGAGGAGCAGTACAAC  
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCTGCACCAGGACTGGCTG  
AATGGCAAGGAGTACAAGTGAAGGTCTCAACAAAGCCCTCCAACCCCC  
ATCGAGAAAACCATCTCAAAGCCAAGGGCAGCCCCGAGAACCCACAGGT  
GTACACCCCTGCCCATCCGGATGAGCTGACCAAGAACCAAGGT CAGCCT  
30 GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATGCCGTGGAGTGGGA  
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCCTCCGTGCTGG  
ACTCCGACGGCTCTCTACAGCAAGCTCACCGTGGACAAGAGCA  
GGTGGCAGCAGGGAAACGTCTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGCAGAAGAGCCTCTCCGTCTCCGGTAAATGAGTGC  
35 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

**Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., *Nature* 256:495 (1975); Köhler et al., *Eur. J. Immunol.* 6:511 (1976); Köhler et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling et al., in: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (*Gastroenterology* 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide.

Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulian et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

**20      Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at  $2 \times 10^5$  cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in  
5 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of  
10 transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off  
15 PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl<sub>2</sub> (anhyd); 0.00130 mg/L

20 CuSO<sub>4</sub>·5H<sub>2</sub>O; 0.050 mg/L of Fe(NO<sub>3</sub>)<sub>3</sub>·9H<sub>2</sub>O; 0.417 mg/L of FeSO<sub>4</sub>·7H<sub>2</sub>O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl<sub>2</sub>; 48.84 mg/L of MgSO<sub>4</sub>; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO<sub>3</sub>; 62.50 mg/L of NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O; 71.02 mg/L of Na<sub>2</sub>HPO<sub>4</sub>; .4320 mg/L of ZnSO<sub>4</sub>·7H<sub>2</sub>O; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic  
25 Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L-Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H<sub>2</sub>O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-  
30 2HCL-H<sub>2</sub>O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H<sub>2</sub>O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22  
35 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H<sub>2</sub>O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of

- Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B<sub>12</sub>; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine;
- 5 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x
- 10 penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

20 It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

#### Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

30 GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in

many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

5       The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

10      The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and  
15     (b) Class 2 includes IFN- $\alpha$ , IFN- $\gamma$ , and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

20      Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

25      Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u>			<u>STATs</u>	<u>GAS(elements) or ISRE</u>
			<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>		
<u>IFN family</u>							
5	IFN-a/B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	Il-10	+	?	?	-	1,3	
<u>gp130 family</u>							
10	IL-6 (Pleiotrophic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	Il-11(Pleiotrophic)	?	+	?	?	1,3	
	OnM(Pleiotrophic)	?	+	+	?	1,3	
	LIF(Pleiotrophic)	?	+	+	?	1,3	
15	CNTF(Pleiotrophic)	-/+	+	+	?	1,3	
	G-CSF(Pleiotrophic)	?	+	?	?	1,3	
	IL-12(Pleiotrophic)	+	-	+	+	1,3	
<u>g-C family</u>							
20	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
25	IL-15	?	+	?	+	5	GAS
<u>gp140 family</u>							
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	5	GAS
<u>Growth hormone family</u>							
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
35	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
<u>Receptor Tyrosine Kinases</u>							
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
40	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to

- 5 bind STATs upon induction with a range of cytokines (Rothman et al., *Immunity* 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTCCCCGAAATCTAGATTCCCCGAAATGATTCCCCG

- 10 AAATGATTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

- 15 PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTCCCCGAAATCTAGATTCCCCGAAATGATTCCCCGAAATG  
20 ATTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCC  
CTAACCTCCGCCATCCGCCCTAACTCCGCCAGTCCGCCATTCTCCGC  
CCCATGGCTGACTAATTTTTATTATGCAGAGGCCGAGGCCCTCGGC  
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTGGAGGCCTAGGCTT  
TGCAAAAAGCTT:3' (SEQ ID NO:5)

- 25 With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase,
- 30 alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

- 35 The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

- 5 Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules 10 containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, Il-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter 15 construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

**Example 13: High-Throughput Screening Assay for T-cell Activity.**

- The following protocol is used to assess T-cell activity by identifying factors, 20 such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and 25 Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

- Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 30 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

- Specifically, the following protocol will yield sufficient cells for 75 wells 35 containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

**Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfet U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^7$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

- 10 15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then  
20 resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

- 25 These cells are tested by harvesting  $1 \times 10^8$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma  
30 can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

**Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, 5 EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or 10 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

15 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCGG -3' (SEQ ID NO:6)  
5' GCAGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

20 Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

25 To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) 30 containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

35 Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

- To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS
- 5 (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

- The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$
- 10 cells/ml.

- Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold
- 15 induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### Example 16: High-Throughput Screening Assay for T-cell Activity

- NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety
- 20 of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and
- 25 antimicrobial responses, and multiple stress responses.

- In non-stimulated conditions, NF-  $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I-  $\kappa$ B is phosphorylated and degraded, causing NF-  $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF-  $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and
- 30 class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating

diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTCCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:  
5' GCGGCCTCGAGGGACTTCCCCGGGGACTTCCGGGGACTTCCGGGAC  
TTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:  
10 5' GCGGCAAGCTTTGCCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)  
15 Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5' CTCGAGGGACTTCCCCGGGGACTTCCGGGGACTTCCGGGACTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCATACTCCGCCCTAACTCCGCCA  
20 TCCCGCCCTAACTCCGCCAGTCCGCCATTCTCCGCCCATGGCTGACT  
AATTTTTTATTATGCAGAGGCCGAGGCCCTCGGCCCTGAGCTATT  
CAGAAGTAGTGAGGAGGCTTTGGAGGCCTAGGCTTGCAAAAGCTT:  
3' (SEQ ID NO:10)

25 Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.  
30

Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

**Example 17: Assay for SEAP Activity**

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

25

**Reaction Buffer Formulation:**

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Bioteck washer with HBSS leaving 100 ul of buffer.

- 5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

- 10 For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

- For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.
- To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca<sup>++</sup> concentration.

20

**Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity**

- The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

- Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

5 Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodynne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr  
10 with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of  
15 alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodynne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of  
20 Loprodynne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na<sub>3</sub>VO<sub>4</sub>, 2 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>  
25 and a cocktail of protease inhibitors (# 1836170) obtained from Boehlerger Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum  
30 manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.  
35 Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

5       The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg<sub>2+</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the  
10 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

15      Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-  
20     POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

25      Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

**Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,  
30     Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other  
35     kinase.

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then 5 rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C 10 until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyn filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts 15 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and 20 Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

25

**Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from 30 these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

35 PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTHERM Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

- PCR products is cloned into T-tailed vectors as described in Holton, T.A. and  
5 Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated  
10 according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

15 Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv.  
20 et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated  
25 disease.

**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample,  
30 and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a  
35 sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

5 The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

10 Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

15 Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale).  
Interpolate the concentration of the polypeptide in the sample using the standard curve.

#### Example 23: Formulating a Polypeptide

20 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

25 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If 30 given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending 35 on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes 5 of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

- 10 Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric 15 acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; 20 U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

- For parenteral administration, in one embodiment, the secreted polypeptide is 25 formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are 30 known to be deleterious to polypeptides.

- Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood 35 of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as 5 ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, 10 manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

15 The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

20 Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

25 Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

30 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

**Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

- 5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily  
10 dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

- 15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5,  
20 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

**Example 26: Method of Treatment Using Gene Therapy**

- 25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is  
30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

5 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

10 The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to 15 transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

20 The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

25 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is 30 required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

- 5        The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

## (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Human Genome Sciences, Inc. et al.  
(ii) TITLE OF INVENTION: 87 Human Secreted Proteins  
(iii) NUMBER OF SEQUENCES: 323  
(iv) CORRESPONDENCE ADDRESS:

10 (A) ADDRESSEE: Human Genome Sciences, Inc.  
(B) STREET: 9410 Key West Avenue  
(C) CITY: Rockville  
(D) STATE: Maryland  
(E) COUNTRY: USA  
(F) ZIP: 20850

## 15 (v) COMPUTER READABLE FORM:

- 20 (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
(B) COMPUTER: HP Vectra 486/33  
(C) OPERATING SYSTEM: MSDOS version 6.2  
(D) SOFTWARE: ASCII Text

## 25 (vi) CURRENT APPLICATION DATA:

- 25 (A) APPLICATION NUMBER:  
(B) FILING DATE: March 19, 1998  
(C) CLASSIFICATION:

## 30 (vii) PRIOR APPLICATION DATA:

- 30 (A) APPLICATION NUMBER:  
(B) FILING DATE:

## 35 (viii) ATTORNEY/AGENT INFORMATION:

- 35 (A) NAME: A. Anders Brookes  
(B) REGISTRATION NUMBER: 36,373  
(C) REFERENCE/DOCKET NUMBER: PZ004PCT

## 40 (vi) TELECOMMUNICATION INFORMATION:

- 40 (A) TELEPHONE: (301) 309-8504  
(B) TELEFAX: (301) 309-8439

## 45 (2) INFORMATION FOR SEQ ID NO: 1:

- 50 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 733 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

	GGGATCCOGGA GCCCAAATCT TCTGACAAAAA CTCACACATG CCCACCGTGC CCAGCACCTG	60
	AATTCGAGGG TGCACCGTCA GTCTTCCCTCT TCCCCCCTAA ACCCAAGGAC ACCCTCATGA	120
5	TCTCCCGGAC TCCTGAGGTG ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG	180
	TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCCGGGG	240
10	AGGAGCAGTA CAACAGCAGC TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAAGACT	300
	GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCA ACCCCCACATCG	360
	AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCCGAGAACCC ACAGGTGTAC ACCCTGCC	420
15	CATCCCGGGA TGAGCTGACC AAGAACCAAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT	480
	ATCCAAGCGA CATGCCGTG GAGTGGGAGA GCAATGGCA GCCGGAGAAC AACTACAAGA	540
20	CCACGCCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG	600
	ACAAGAGCAG GTGGCAGCAG GGGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC	660
	ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC	720
25	GACTCTAGAG GAT	733

## 30 (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp	Ser	Xaa	Trp	Ser
40	1		5	

## 45 (2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 86 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

55	GCGCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTTCCCG AAATGATTTC	60
	CCCGAAATAT CTGCCATCTC AATTAG	86

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGGCAAGCT TTTTGCAAAG CCTAGGC

27

15

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 271 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATTT CCCCCGAAATC TAGATTCCCG CGAAATGATT TCCCCGAAAT GATTTCCCCG

60

AAATATCTGC CATCTCAATT AGTCAGCAAC CATACTCCCG CCCCTAACTC CGCCCATCCC

120

30 GCCCCCTAACT CCGCCCAGTT CCGCCCATTC TCCGCCCAT GGCTGACTAA TTTTTTTTAT

180

TTATGCAGAG GCCGAGGCCG CCTCGGCCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT

240

35 TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T

271

40 (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

50 GCGCTCGAGG GATGACAGCG ATAGAACCCC GG

32

55 (2) INFORMATION FOR SEQ ID NO: 7:

## (i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

5

GCGAAGCTTC GCGACTCCCC GGATCCGCCT C

31

10

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

20

GGGGACTTTC CC

12

25

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

35

GCGGCCTCGA GGGGACTTTC CGGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG

60

CCATCTCAAT TAG

73

40

(2) INFORMATION FOR SEQ ID NO: 10:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTTCCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT

60

55

CAATTAGTCA GCAACCATAG TCCCGCCCCCT AACTCCGCC ATCCCGCCCC TAACTCCGCC

120

CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA

180

60

GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTG GAGGCCCTAGG

240

CTTTTGCAAA AAGCTT

256

5

(2) INFORMATION FOR SEQ ID NO: 11:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1679 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

15	GCAGCGCAC CGGGCGATCG CTTCACGGAT GCGGACGACG TAGCCATCCT TACCTACGTG	60
20	AAGGAAAATG CCCGCTCGCC CAGCTCCGTC ACCGGTAACG CCTTGTGGAA AGCGATGGAG	120
25	AAGAGCTCGC TCACGCCAGCA CTCTGCGAG TCCCTGAAGG ACCGCTACCT CAAGCACCTG CGGGGCCAGG AGCATAAGTA CCTGCTGGGG GACGCCCGG TGAGCCCCTC CTCCCAGAAG CTCAGCGGA AGGGGGAGGA GGACCCGGAG GCGGGGGATA GCGGGGAACC ACAGAATAAG	180 240 300
30	AGAAACTCCAG ATTTGCCTGA AGAAGAGTAT GTGAAGGAAG AAATCCAGGA GAATGAAGAA GCAGTCAAAA AGATGCTTGT GGAAGCCACC CGGGAGTTTG AGGAGGTTGT GGTGGATGAG	360 420
35	AGCCCTCCTG ATTTTGAAAT ACATATAACT ATGTGTGATG ATGATCCACC CACACCTGAG GAAGACTCAG AAACACAGCC TGATGAGGAG GAAGAAGAAC AGAAAGAAAA AGTTTCTCAA CCAGAGGTGG GAGCTGCCAT TAAGATCATT CGGCAGTTAA TGGAGAAGTT TAACTTGGAT	480 540 600
40	CTATCAACAG TTACACAGGC CTTCTAAAAA AATAGTGGTG AGCTGGAGGC TACTTCCGCC TTCTTAGCGT CTGGTCAGAG AGCTGATGGA TATCCATTG GGTCCCGACA AGATGACATA GATTTGCAAA AAGATGATGA GGATACCAAGA GAGGCATTGG TCAAAAAATT TGGTGCTCAG AATGTAGCTC GGAGGATTGA ATTCGAAAG AAATAATTGG CAAGATAATG AGAAAAGAAC AAAGTCATGG TAGGTGAGGT GGTAAAAAAA AATTTGACC AATGAACCTT AGAGAGTTCT	660 720 780 840 900
45	TGCATTGGAA CTGGCACTTA TTTTCTGACC ATCGCTGCTG TTGCTCTGTG AGTCCTAGAT TTTTGTAGCC AACCAAGAGTT GTAGAGGGGG ATAAAAAGAA AAGAAATTGG ATGTATTTAC AGCTGTCTT GAACAAGTAT CAATGTGTTT ATGAAAGGAA GATCTAAATC AGACAGGAGT TGGTCTACAT AGTAGTAATC CATTGTTGGA ATGGAACCCCT TGCTATAGTA GTGACAAAGT GAAAGGAAAT TTAGGAGGCA TAGGCCATT CAGGCAGCAT AAGTAATCTC CTGTCCTTGT	960 1020 1080 1140 1200
50	GCAGAAGCTC CTTTAGATTG GGATAGATTG CAAATAAAGA ATCTAGAAAT AGGAGAAGAT TTAATTATGA GGCTTGAAAC ACGGATTATC CCCAAACCTC TGTCATTCC CCCAGTGAGC TCTGATTCT AGACTGCTT GAAAATGCTG TATTCTTTT GCTAACTTAG TATTGGGTA	1260 1320 1380

	CCCTGCTCTT TGGCTGTTCT TTTTTGGAG CCCCTCTCAG TCAAGTCAGC CGGATGCTTT	1440
5	TCTTITACCTA CCCCTCAGTT TTCCCTAAAA CGCGCACACA ACTCTAGAGA GTGTTAAGAA	1500
	TAATGTTACT TGGTTAATGT GTTATTATT GAGTATTGTT TGTGCTAAGC ATTGTGTTAG	1560
	ATTTAAAAAA TTAGTGGATT GACTCCACTT TGTTGTGTTG TTTTCATTTG TGAAAATAAA	1620
10	TATAACTTTG TATTCGAAAA AAAAAAAAAA AAAATNRCTG CGGNCCGACA AGGGAATT	1679

## 15 (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1830 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25	GCGACCGCGC CCTTCAGCTA GCTCGCTCGC TCGCTCTGCT TCCCTGCTGC CGGCTGCGCA	60
	TGGCTTNGGC GTTGGCGGGCG CTGGCGGGGG TCGAGCNGCC TCGGSAGCCG GTACCAGCAG	120
30	TTGCAGAACATG AAGAAGAGTC TGGAGAACCT GAACAGGCTG CAGGTGATGC TCCCTCACCT	180
	TACAGCAGCA TTTCTGCAGA GAGCGCACAT NATTTGACT ACAAGGATGA GTCTGGGTTT	240
	CCAAAGCCCC CATCTTACAA TGTAGCTACA ACACTGCCA GTTATGATGA AGCGGAGAGG	300
35	ACCAAGGCTG AAGCTACTAT CCCTTTGGTT CCTGGGAGAG ATGAGGATTT TGTGGGTCGG	360
	GATGATTTG ATGATGCTGA CCAGCTGAGG ATAGGAAATG ATGGGATTTT CATGTTAACT	420
40	TTTTTCATGG CATTCTCTT TAACTGGATT GGGTTTTCC TGTCTTTTG CCTGACCACT	480
	TCAGCTGCAG GAAGGTATGG GGCCATTTCA GGATTTGGTC TCTCTCTAAT TAAATGGATC	540
	CTGATTGTCA GGTTTTCCAC CTATTTCCCT GGATATTTTG ATGGTCAGTA CTGGCTCTGG	600
45	TGGGTGTTCC TTGTTTTAGG CTTCTCCTG TTTCTCAGAG GATTTATCAA TTATGCAAAA	660
	GTTCGGAAGA TGCCAGAAC TTTCTCAAAT CTCCCCAGGA CCAGAGTTCT CTTTATTTAT	720
50	TAAAGATGTT TTCTGGCAA GGCCCTCCTG CATTATGAA TTCTCTCTCA AGAAGCAAGA	780
	GAACACCTGC AGGAAGTGAA TCAAGATGCA GAACACAGAG GAATAATCAC CTGCTTTAAA	840
	AAAATAAAAGT ACTGTTGAAA AGATCATTTC TCTCTATTG TTCCTAGGTG TAAAATTTA	900
55	ATAGTTAATG CAGAATTCTG TAATCATTGA ATCATTAGTG GTTAATGTTT GAAAAAGCTC	960
	TTGCAATCAA GTCTGTGATG TATTAATAAT GCCTTATATA TTGTTTGTAG TCATTTAAG	1020
60	TAGCATGAGC CATGTCCCTG TAGTCGGTAG GGGGCAGTCT TGCTTTATTIC ATCCTCCATC	1080

	TCAAAATGAA CTTGGAATTA AATATTGTAA GATAATGTATA ATGCTGGCCA TTTTAAAGGG	1140
	GTTTTCTCAA AAGTTAAACT TTGTTATGA CTGTGTTTT GCACATAATC CATATTGCT	1200
5	GTTCAAGTTA ATCTAGAAAT TTATTCAATT CTGTATGAAC ACCTGGAAGC AAAATCATAG	1260
	TGCAAAAATA CATTAAAGGT GTGGTCAAA ATAAGTCTTT AATTGGTAAA TAATAAGCAT	1320
10	TAATTTTTA TAGCCTGAT TCACAATTCT GCGGTACCTT ATTGTACCTA AGGGATTCTA	1380
	AAGGTGTTGT CACTGTATAA AACAGAAAGC ACTAGGATAC AAATGAAGCT TAATTACTAA	1440
	AATGTAATTCT TTGACACTCT TTCTATAATT AGCGTTCTTC ACCCCCACCC CCACCCCCAC	1500
15	CCCCCTTATT TTCCCTTTGT CTCCTGGTGA TTAGGCCAAA GTCTGGGAGT AAGGAGAGGA	1560
	TTAGGTACTT AGGAGCAAAG AAAGAAGTAG CTTGGAACCTT TTGAGATGAT CCCTAACATA	1620
20	CTGTACTACT TGCTTTTACA ATGTGTTAGC AGAAACCAGT GGGTTATAAT GTAGAATGAT	1680
	GTGCTTTCTG CCCAAGTGTT AATTCACTTT GGTTTGCTAT GTAAAAACTG TAAATACAAC	1740
	AGAACATTAATCTC TTGTGTAGCA CCTTTTAAAAA AAAAAAAAAA AAAAAAAAAA	1800
25	AAAAAAAAAA AANCCCGGGG GGGGGCCCN	1830

## 30 (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1212 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

40	TGTTTGAAGT TGTTACTTTT GTTTACAGCA AAGTTTGATG TAGTGTGCAG TAGTGAGCTC	60
	TAGACTGATC TTTTTCTAAA TCAGAAAGTG ATTAAAGTAT GCACAACCAA AGGCAGGTTT	120
	TTCTTTTCA TTTATTCAAGC AACTATTATTA TAAGCATCAA CTCTGTGCCA GGCACGTTAC	180
45	TAGCTGCTAC ATACTGCTG AACATGACAT ACGGTTAAGT AACTTTACAA TTATTATCAA	240
	ATACTTCAAT GTAGATATTCTTAAAGTTGA AATAGCATTA ACTAGGATAA TGCTTTCATG	300
50	TTATTTTATT TGTCTTGTGA TAGAAATTCA ACTTTGTACC ATCTTAAAC TAGGTTGCTA	360
	AAAAAAATAGG AGGATGAAGT CAATAAAGTT TATGCCAGTT TAAAAACTGG AAGGAAAAGG	420
	TAAGAGCTCT CCATTATAAA ATAGTTGCAT TCGGTTAATT TTTACACATT AGTGCATTGC	480
55	GTATATCAAC TGGCCCTCAA TGAAGCATT AAGTGCTTGG AATTTTACTA AACTGACTTT	540
	TTTGCAACTT TGGGAGATT TTGAGGGAG TGTGAAAAT TGCCAAACAC TCACCTCTTA	600
60	CTCAAAACTT CAAATAAAAT ACACATTTC AAGAGGGAGC ACCTTTTATA TTTGATAAGT	660

	TTTCATTATA AACCTTATAA TACCA GTCAC AAAGAGGTTG TCTGTCTATG GTTTAGCAA	720
5	CATTTGCTTT CTCTTTGGA AGTGTGATTG CAATTGCAGA ACAGAAAGTG AGAAAACACT	780
	GCCAGCGGTG ATTGCTACTT GAGGTAGTTT TTTACAACTA CCATTTCCCC TCCATGAAAT	840
	TATGTGAAAT TTATTTTATC TTGGGAAAA GTTGAGAAGA TAGTAAAGA ATTTAGGAATT	900
10	TAAAATTACA GGGAAAAATA TGTAAGTGAA AAGCAATAAA TATTTTGTTC ACTTTGCTAT	960
	CAAGATGTTT ACTATCAGAT ATTTATTATA TGGCAGCAAT TTATATTTTT AATCATGCC	1020
15	CATTAATAGA CGCAGTAAAA TATTTTGAA TCAGACATTG GGGGTTGTA TGTGCATTAA	1080
	AATTGTCCTT TGTACTGTAA GTTACTGTTA ATTTGAATAT TTATATTGAAAC TGTCTCCCTG	1140
	TGCCTTTATA ATATAAAGTT GTTCTACAA CTTTTAATGA TCTTAATAAA GAATACTTTA	1200
20	AGAAAAAAA AA	1212

## 25 (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2061 base pairs
- (B) TYPE: nucleic acid
- 30 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

35	GGTTTTCCTC CGACTTCCGG ACATCTCCCT GGGAGTCGG CAGAGTGGAG TCAAAGGCAA	60
	CCAGTGCTCG CTGCGGTCTC TGGGGATCGG GACCGCGCG GCAGGCCCG AGCGGGATGT	120
40	TCCGGGGCTT GAGCAGTTGG TTGGGCTTGC AGCAGCCGT GGCAGGCCGT GGCGAGCCC	180
	ATGGAGATGC TCCACCCGAG CAGCCGTCCG AGACGGTGGC TGAGTCTGCG GAGGAGGAGC	240
	TGCAGCAAGC GGGAGACCAG GAGCTCCTCC ACCAGGCCAA AGACTTCGGC AACTATTTAT	300
45	TTAACTTTGC ATCTGCTGCC ACAAAAAAGA TAACTGAATC AGTTGCTGAA ACAGCACAAA	360
	CAATAAAGAA ATCCGTAGAA GAAGGAAAAA TAGATGGCAT CATTGACAAG ACAATTATAG	420
50	GAGATTTCA GAAGGAACAG AAAAAATTG TTGAAGAGCA ACATACAAAG AAGTCAGAAG	480
	CAGCTGTGCC CCCATGGTT GACACTAACG ATGAAGAAC AATTCAACAA CAAATTTGG	540
	CCTTATCAGC TGACAAGAGG AATTTCCTTC GTGACCCCTCC GGCTGGCGTG CAATTTAATT	600
55	TCGACTTTGA TCAGATGTAC CCCGTGGCCC TGGTCATGCT CCAGGAGGAT GAGCTGCTAR	660
	CAAGATGAGA TTTGCCCTCG TTCCTAAACT TGTGAAGGAA GAAGTGTCT GGAGGAAC	720
60	CTTTTACCGC GTCTCCCTGA TTAAGCAGTC AGCCAGCTC ACGGCCCTGG CTGCCAAC	780

	GCAGGCCGCA GGGAAAGGGAG GAGAAGAGCA ATGGCAGAGA GCAAGATTTG CCGCTGGAGA	840
	GGCAGTACGG CCCAAAACGC CACCCGTTGT AATCAAATCT CAGCTTAAAA CTCAAGAGGA	900
5	TGAGGAAGAA ATTTCTACTA GCCCAGGTGT TTCTGAGTTT GTCACTGATG CCTTCGATGC	960
	CTGTAACCTA AATCAGGAAG ATCTAAGGAA AGAAATGGAG CAACTAGTGC TTGACAAAAA	1020
10	GCAAGAGGAG ACAGCCGTAC TGGAAGAGGA TTCTGCAGAT TGGGAAAAAG AACTGCAGCA	1080
	GGAACCTCAA GAATATGAAG TGGTGACAGA ATCTGAAAAA CGAGATGAAA ACTGGGATAA	1140
	GGAAATAGAG AAAATGCTTC AAGAGGAAAA TTAGCTGTTCTGAAATAGA AGAATAATCC	1200
15	TTAACACTCT GCAAACTGAC ATTAATTCT AGATGTTGAC AATTACTGAA TCAGAAGGCA	1260
	TGAAAGAGTA TAATTTTATG AAATTCAAAA TTATTCTTTT TTCAAGTTGA AACTTGCTC	1320
20	TTCTACTTTA AAAAGTATA TAGAACAGTT ACTTCTAATA ATCAGAAAGA GATGTTTAT	1380
	AGAACATTTTC TTTAATATAA AGTTAGAGAT GTCTTCATAG GCAGTATGGC TATCTTGCC	1440
	ACAGAACAT AAGTAAAATT TTAGAGTTCT GTTTTCCATG AGGTCAAAAA TATAATTAT	1500
25	TCCTCAGTCA TGGTTTCTA AATATCTGTA CTCCACATTC CATTAAATT GATATGAGGG	1560
	TGTTAAAGTA CCTACTTAAT GGGTGATTA CTATCAAAAT GACCAAATTA TACCAAAGAA	1620
30	CTTAAGAGGA AGCACTTCA GAACTATTCA CTTGCCAGGT ATTTCTAAA ATTCCACCTG	1680
	AAAGCCAAAA GATAAAATAC ATNAGTTGGA TTTTAATGAT ATAAGCATCA CACAATTITA	1740
	CATTAAGAAA TACTGTGCAG CCCATGCGTG GTGGCTCAGG CCTGTAATCC CAGCANTTG	1800
35	GGAGGCCGAG GTGGGCAGAT CACCGGAGGT CAGGAGTTCG AGACCAGCCT TGCCAACATA	1860
	GTGAAACCT GTCTTTACTA AAAATACAAA AATTAGCCGG GCATGGTGGC AGGCACCTGT	1920
40	AATCCCAGCT ACTAGGGAGG CTTTGAAACC CAGGAGGCAG AGGTGAGCAGC GAGCTGAGAT	1980
	CGCGCCACTG CACTCCAGCC TGGGTGATAG AGTGAGATTC AGTCTCAAAA AAAAAAAAAAA	2040
	AAAAAAAAAAA AATGACCTCG A	2061
45		

## (2) INFORMATION FOR SEQ ID NO: 15:

50	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1412 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
55	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:	
	CCCTTCATCT GCGTTGCCAG GAACCTGTGTC AGCAGAAACT TCTCAAGCCC CATCCTTGCC	60
60	AGGAAGCTCT GTGAAGGTGC TGCTGATGAC CCAGATTCT CCATGGTCCT CCTGTGTC	120

	CTGTTGGTGC CCCTCCCTGCT CAGTCTCTTT GTACTGGGGC TATTTCTTTG GTTTCCTGAAG	180
5	AGAGAGAGAC AAGAAGAGTA CATTGAAGAG AAGAAGAGAG TGGACATTTG TCGGGAAACT	240
	CCTAACATAT GCCCCCCTTC TGAGAGAAC ACAGAGTACG ACACAATCCC TCACACTAAT	300
	AGAACAAATCC TAAAGGAAGA TCCAGCAAAT ACCGTTTACT CCACGTGGA AATACCGAAA	360
10	AAGATGGAAA ATCCCCACTC ACTGCTCACG ATGCCAGACA CACCAAGGCT ATTTGCCTAT	420
	GAGAATGTGA TCTAGACAGC AGTGCCTC CCTAAGTCTC TGCTCAAAAA AAAAACAAATT	480
15	CTCGGCCAA AGAAAACAAT CAGAAGAATT CACTGATTG ACTAGAAACA TCAAGGAAGA	540
	ATGAAGAACG TTGACTTTT TCCAGGATAA ATTATCTCTG ATGCTCTTT AGATTTAAGA	600
	GTTCATTAATT CCATCCACTG CTGAGAAATC TCCTCAAACC CAGAAGGTTT AATCACTTCA	660
20	TCCCAAAAT GGGATTGTGA ATGTCAGCAA ACCATAAAAA AAGTGCTTAG AAGTATTCCCT	720
	ATAAAAATGT AAATGCAAGG TCACACATAT TAATGACAGC CTGTTGTATT AATGATGGCT	780
25	CCAGGTCACT GTCTGGAGTT TCATTCATC CCAGGGCTTG GATGTCAGGA TTATACCAAG	840
	AGTCTTGCTA CCAGGAGGGC AAGAAGACCA AAACAGACAG ACAAGTCCAG CAGAAGCAGA	900
	TGCACCTGAC AAAATGGAT GTATTAATTG GCTCTATAAA CTATGTGCC AGCAYTATGC	960
30	TGAGCTTACA CTAATTGGTC AGACATGCTG TCTGCCCTCA TGAAATTGGC TCCAAATGAW	1020
	TGAACTACTT TCATGAGCAG TTGTAGCAGG CCTGACCACA GATTCCCAGA GGGCCAGGTG	1080
35	TGGATCCACA GGACTTGAAG GTCAAAGTTC ACAAAAGATGA AGAACATCAGGG TAGCTGACCA	1140
	TGTTTGGCAG ATACTATAAT CGAGACACAG AAGTGTCCAT GGCCCAAGGA CAAGGACCTC	1200
	CAGCCAGGCT TCATTTATGC ACTTGTCTGC AAAAGAAAAG TCTAGGTTTT AAGGCTGTG	1260
40	CAGAACCCAT CCCAATAAAG AGACCGAGTC TGAAGTCACA TTGAAATCT AGTGTAGGAG	1320
	ACTTGGAGTC AGGCAGTGAG ACTGGTGGGG CACGGGGGGC ANTGGGTANT GTAAACCTTT	1380
45	TAAAGATGGT TAATTCTCA TTAGTGTGTT TT	1412

50 (2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1052 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

60 TTCCCTCTCCT CTCTCTACCC CTCCGTCTC TCCTCCCCCTC CTCTCTCTTC CTCTCCTCTC 60

	TCTCTTCCCTC TCCTCTCTCT TCCCTTCCTG TCTCTCTTCC CCTCCTCTCT CTCTTCCGT	120
	CCTCTATCTC TTCCCCCTCT CTATCTCTC CTCTCCCTCTC TCTCTTCCCTC CCCTCTCTCT	180
5	CTCTTSCPTT CTCTCTCTC TCCTGTCTCG GCTGTTGTT GGTCAGGTT GGGTGCCTGCT	240
	GTTGTGGTCC TTCCCAGAAA CTGCCAGTAG AGGGCAGCCT GGGCATCCTA ATGCTTACTC	300
10	TGGTTGTTAC ACAAAAGAAA TATTGGGGTC ACTGGCGAGC CCACCCACAC TCACCAGAAT	360
	CTCCACTGTA GTCCCCCTAA CAAACAGCCC TTCACTTCCT CTCCCCACTTC AGCAATTGT	420
	ATTTTGATGC CATTGGCCTC AGATCAGAGT GTTTTAAATC ATCACGCCCT GGCTTATCCC	480
15	TGGTCGAGCC AGGACACGGG GTGCTTCAGT GGGTCTGTCA CCCTCTCTCC TTGAAGCATG	540
	TTGCTTTAT TTATTTACTT TTACTCTCAC CCTGCTCTG TACCAGCAGG GGCCACTTCA	600
20	AAGCCAAGGT ACAGGGTGAT AACTTGTGGT CCAGCCTCAG TTTTCTCCAC TTCTTTCTCC	660
	CACTCACCCC CAGCAAGGTG CCTGGGGAGA CTTGAGCAGA TGTTTCATTG TGGCTGGCC	720
	AGTGGCTGAA AGCAGGCCTC CAATGCACTG TGACCTCTGG CTTCCCCAGC AGCTTTCCCA	780
25	GAGAGGCAGA GGGGCCTTCC ACAGCCCCGGG TTCTCCTGCT GCCTCCTGCC TGCTGCAGCT	840
	GCAGGCATTC TGAGGGCAA CGTGGAGGAA GGGCCAGGGA TGCAATGGAT TTTAATTGTT	900
30	TCATCACACC TTCCCCGTGG CAAAGAAACA GTCAGTCCTC TTCAGGTGTC TTCTGGATTG	960
	CTGGTGATGG ACAGAGAAAT CTTTTTACAG TTTCAAATTA TGTTCAACAA ATAAAAAATTG	1020
	CATTTTTAT TTGGAAAAAA AAAAAAAAAA AA	1052

35

## (2) INFORMATION FOR SEQ ID NO: 17:

## 40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 683 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

	AATTCCGGCAG AGGCCATTAT CATGTACATA TAGCCTGTTT TTTACCATG TGAGACAAAG	60
50	TAGGCATATT CCTTTCCATC CAAGAACTCA TAACCTAGTA ATTGTAGTTG GCTGATAGCT	120
	CATTGCCAT ACACAAGGAT CTAACACAAAC CTCTTGAATA AACATCCCCC TTATTCAAGAA	180
	ATGCCTTTTC CTATTTCCAT ATTGCAACTT TGCTTACAAA TTTCCAATCT GTCTTTCTGT	240
55	TTACAGAAGA TATACAAAAT TCCCTTTGTA TGATCTCTTT ATATCTCTG ATTTCTTTT	300
	GTGTTTGCTA CCAAAGGGCC TGCACATAGT GAGAAGATTG TGCAATGATCT GTGAGCTCTA	360
60	CCACACCTGG AATTAGGGAT CACCAATATG AGAAAAAAAT TTGGAGGTAC AAATAACATT	420

	ATCATATGTW ATTGGCATAT AAATTACAGA TGTWICATATG ACTAAAAACC CTGTGGATAT	480
5	WAACCMATG CAGATAANIW TAATAAAATW TWTAATAATW TWATCMAATA ATGATAGTGC	540
	TATTCAAATA CTTCAAATTT GCACAGTGAT TTATTCITTA AAATATGTTA ACACATGTGA	600
	GCCAATACAC TGAGGTCACT GGATAAATAA ACAGATTCTT GCAAAAAAAA AAAAAAAA	660
10	ACTCGAGGGG GGCCCGTACC CTT	683

## 15 (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1054 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

25	AAACTCATTT AGGTGACACT ATAGAAGGTA CGCCTGCAGG TACCGGTCCG GAATTCCCGG	60
	GTCGACCCAC GMGNCCGGCG ACAAGATGGC AGCAGCGTGT CGGAGCGTGA AGGGCCTGGT	120
30	GGCGGTAAATA ACCGGAGGAG CCTCGGGCCT GGGCTGGCC ACGGCGGACG ACTTGTGGGG	180
	CAGGGAGCCT CTGCTGTGCT TCTGGACCTG CCCAACTCGG GTGGGGAGGC CCAAGCCAAG	240
	AAGTTAGGAA ACAACTGCGT TTTCGCCCCA GCCGACGTGA CCTCTGAGAA GGATGTGCAA	300
35	ACAGCTCTGG CTCTAGAAA AGGAAAGTTT GGCGTGTGG ATGTAGCTGT CAACTGTGCA	360
	GGCATCGCGG TGGCTAGCAA GACGTACAAC TTAAAGAAGG GCCAGACCCA TACCTTGAA	420
40	GACTTCCAGC GAGTTCTTGA TGTGAATCTC ATGGCACCT TCAATGTGAT CCGCCTGGTG	480
	GCTGGTGAGA TGGGCCAGAA TGAACCAGAC CAGGGAGGCC AACGTGGGGT CATCATCAAC	540
	ACTGCCAGTG TGGCTGCCCT CGAGGGTCAG GTGGACAAG CTGCATACTC TGCTTCCAAG	600
45	GGGGGAATAG TGGGCATGAC ACTGCCATT GCTCGGGATC TGGCTCCCAT AGGTATCCGG	660
	GTGATGACCA TTGCCCCAGG TCTGTTGGC ACCCCACTGC TGACCAGCCT CCCAGAGAAA	720
50	GTGTGCAACT TCTTGGCCAG CCAAGTCCC TTCCCTAGCC GACTGGGTGA CCCTGCTGAG	780
	TATGCTCACC TCGTACAGGC CATCATCGAG AACCCATTCC TCAATGGAGA GGTCACTCCG	840
	CTGGATGGGG CCATTCTGAT GCAGCCTTGA AGGGAGAAGG CAGAGAAAAC ACACGCTCCT	900
55	CTGCCCTTCC TTTCCCTGGG GTACTACTCT CCAGCTTGGG AGGAAGCCCA GTAGCCATT	960
	TGTAACTGCC TACCAAGTCGC CCTCTGTGCC TAATAAGTC TCTTTTCTC ACANAAAAAA	1020
60	AAAAAAAAA AAAAAAAA AAAAAAAA AAAA	1054

## (2) INFORMATION FOR SEQ ID NO: 19:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1393 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

15	GGAACAAAGCT GGGATATGTG AGCGTTAACGC TACTCACATC CTTCAAAAAG GTGAAACATC TTACACGGGA CTGGAGAACCC ACAGGACATG CTTTGAAGTA TTCAAGTGGTC CTTGAGTTGA	60
20	ATGAGGNCCA CGGGAAGGTG AGGAGGACCA CCCCCGTCCC ACTGTTCCCC AACGAGAACCC TCCCCAGCAA GATGCTCCTG GTCTATGATC TCTACTTGTG TCTTAAGCTG TGGGCTCTGG CCACCCCCCA GAAGAATGGG AAGGGTGCCTA GARAAGGTGA TGGAACACCT GCTCAAGCTT	120
25	TTTGGGACTT TTGGAGTCAT CTICATCAGTG CGGATCCTCA AACCTGGGAG AGAGCTGCC CCTGACATCC CGAGGNTCCA GCAGCCGCTA CAGCTCCTCT GACCCCGAGA GCAACCCAC ATCCCCATAG GCGGGCCGAC GGCACGNGKC CACCAACAAG CTCAGCCCGT CTGGCCACCA	180
30	GAATCTCTTT CTGAGTCCAA ATGCCTCCCC GTGCACAAGT CCTTGGAGCA GCCCCTTGGC CCAACGCAAA GGCGTTTCCA GAAAGTCCCC ACTGGCGGAG GAAGGTAGAC TGAACCTGCAG CACCAGCCCT GAGATCTTCC GCAAGTGTAT GGATTATTCC TCTGACAGCA GCGTCACTCC	240
35	CTCTGGCAGC CCCTGGGTCC GGAGGGCTCG CCAAGCCGAG ATGGGGACCC AGGAGAAAAG CCCCGGTACG AGTCCCCCTGC TCTCCCGAA GATGCAGACT GCAGATGGGS TACCCGTAGG	300
40	TNGCTTGAGG TTGCCAGGG GTCCTGACAA CACCAGAGGA TTTCATGGCC ATGAGAGGAG CAGGGCCTGT GTATAAATAC CTTCTATTTT TAATACAAGC TCCACTGAAA ACCACCTTCG	360
45	TTTTCAAGGT TCTGACAAAC ACCTGGCATG ACAGAATGGA ATTGTTCCCC CTTTGAGAGA TTTTTTATTC ATGTAGACCT CTTAATTTAT CTATCTGTAA TATACATAAA TCGGTACGCC	420
50	ATGGTTTGAA GACCACCTTC TAGTTCAAGGA CTCCCTGTCT TCCCAGCATG GCCACTATTT TGATGATGGC TGATGTTGAGT GAGTGTGATG GCCCTGAAGG GCTGTAGGAC GGAGGTTCCC	480
55	TGGGGGAAGT CTGTTCTTGTG GTATGGAATT TTTCTCTCTT CTTTGGTATG GAATTTTCC CTTCAGTGAC TGAGCTGTCC TCGATAGGCC ATGCAAGGGC TTCTTGAGAG TTCAGGAAAG TTCTCTTGAG CAACAGCAAG TAGCTAAGCC TATAGCATGG TGTCTTGAG GACCAATCG	540
60	ATGTTACCTG TCAAGTAAAT AAATAATAAA ACACCCAATC GGGAGTGCTG AAAAAAAANA ANNAAAAAC TCG	600
		1260
		1320
		1380
		1393

## 5 (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1215 base pairs
- (B) TYPE: nucleic acid
- 10 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

15	AGGAAAAGTT TTCCNAATTG GAAAGCGGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG	60
	NTCANTCATT AGGCACCCCA GGCTTACAC TTTATGCTTC CGGNTCGTAT GTTGTGTGGA	120
20	ATTGTGACCG GATAACAATT TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCTN	180
	TAATACGACT CACTATAGGG AAAGCTGGTA CGCCTGCAGG TACCGGTCCG GAATTCCCGG	240
	GTCGACCCAC GCGTCCGCCA ACACGGTCCGT GAAAATCCGA AGTGCAGCGGG AAAGTGGAGG	300
25	TGAGGGCCGC CCGCCCTAGA GGTGCCCGTC CGAGAGGCAG AGCTGACAAG GAAGGTTTCG	360
	AGCGTTTTGC TGGCAAAGGG ATTTCTTACA ACCTCCAGGC ATGCGTCTTT CTGCCCTGCT	420
30	GGCCTTGGCA TCCAAGGTCA CTCTGCCCCC CCATTACCGC TATGGATGA GCCCCCCCAGG	480
	CTCTGTTGCA GACAAGAGGA AGAACCCCCC ATGGATCAGG CGGCGCCCAAG TGGTTGTGGA	540
	ACCCATCTCT GATGAAGACT GGTATCTGTT CTGTGGGAC ACGGTGGAGA TCCTAGAAGG	600
35	CAAGGATGCC GGGAAAGCAGG GCAAAGTGGT TCAAGTTATC CGGCAGCGAA ACTGGGTGGT	660
	CGTGGGAGGG CTGAACACAC ATTACCGCTA CATTGGCAAG ACCATGGATT ACCGGGGAAC	720
40	CATGATCCCT AGTGAAGCCC CCTTGCTCCA CCGCCAGGTC AAACTTGTGG ATCCTATGGA	780
	CAGGAAACCC ACTGAGATCG AGTGGAGATT TACTGAAGCA GGAGAGCGGG TACGAGTCTC	840
	CACACGATCA GGGAGAATTA TCCCTAAACC CGAATTTCCC AGAGCTGATG GCATCGTCCC	900
45	TGAAACGTGG ATTGATGGCC CCAAAGACAC ATCAGTGGAA GATGCTTTAG AAAGAACCTA	960
	TGTGCCCTGT CTAAAGACAC TGCAGGAGGA GGTGATGGAG GCCATGGGA TCAAGGAGAC	1020
50	CCGGAAATAC AAGAAGGTCT ATTGGTATTG AGCCTGGGC AGAGCAGCTC CTCCCCAACT	1080
	TCTGTCCCAG CCTTGAGGC TGAGGCACIT CTTTTTCAGA TGCCAATAAA GAGCACTTTA	1140
	TGAGTCCTCC AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA	1200
55	AAAAGGGCGC GCCGC	1215

## 60 (2) INFORMATION FOR SEQ ID NO: 21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2042 base pairs  
 (B) TYPE: nucleic acid  
 5 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

10	CTGCATCCAG GCGCAGAATA ACCTGGGTAT CTTGTTGCT GAAAGAGAGA AATTGAAACT	60
	GCACAGGCTT ACCTAGAGTC ATCAGAACCA CTATATAATC AGTATATGAA AGAGGTTGGG	120
15	AGTCCTCCCTC TTGATCCTAC TGAGCGTTTT CTTCTGAAGA AGAGAAACTT ACTGAACAAG	180
	AGAGATCAA AAGATTTGAA AAGGTTTATA CTCATAACCT ATATTACCTA GCTCAAGTCT	240
	ACCAGCATCT GGAAATGTTT GAGAAGGCTG CTCACTATTG CCATAGTACA CTAAAACGCC	300
20	AGCTTGAGCA CAATGCCCTAC CATCCTATAG AGTGGGCTAT CAATGCTGCT ACCTTGTAC	360
	AGTTTTACAT CAATAAGCTA TGCTTTATGG AGGCCAGGCA CTGTTTATCA GCTGCTAATG	420
25	TCATTTTTCGG TCAAACCTGGA AAGATCTCAG CCACAGAAGA CACTCCTGAA CCTGAAGGAG	480
	AAGTGCCAGA GCTTTATCAT CAAAGAAAGG GGGAAATAGC AAGGTGCTGG ATCAAATACT	540
	GTTTGACTCT CATGCAGAAT GCCCAACTCT CCATCCAGGA CAACATAGGA GAGCTTGATC	600
30	TTGATAAACCA GTCTGAACCT AGAGCTTTAA GGAAAAAAGA ACTAGATGAG GAGGAAAGCA	660
	TTCGGAAAAA AGCTGTGCAG TTTGGAACCG GTGAACGTG TGATGCCATC TCTGCAGTAG	720
35	AAGAGAAAGT GAGCTACTTG AGACCTTTAG ATTTTGAAGA AGCCAGAGAA CTTTTCTTAT	780
	TGGGTCAAGCA CTATGCTTT GAGGCAAAAG AGTCTTTCA GATTGATGGT TATGTCACTG	840
	ACCATATTGA AGTTGTCCAA GACCACAGTG CTCTGTTAA GGTGCTTGCA TTCTTTGAAA	900
40	CTGACATGGG GAGACGGTGC AAGATGCATA AACGCRGAAT AGCCATGCTA GAGCCCCTAA	960
	CTGTAGACCT GAATCCACAG TATTATCTGT TGGTCAACAG ACAGATCCAG TTTGAAATTG	1020
45	CACATGCTTA CTATGATATG ATGGATTTGA AGGTTGCCAT TGCTGACAGG CTAAGGGATC	1080
	CTGATTCAACA CATTGTAAAA AAAATAAATA ATCTTAATAA GTCAGCACTG AAGTACTACC	1140
	AGCTCTTCTT AGACTCCCTG AGAGACCCAA ATAAAGTATT CCCTGAGCAT ATAGGGGAAG	1200
50	ATGTTCTTCG CCCTGCCATG TTAGCTAAGT TTGAGTTGC CCGTCTCTAT GGCAAAATCA	1260
	TTACTGCAGA TCCCAAGAAA GAGCTGGAAA ATTTGGCAC ACATTTGGGA ACATTACAAA	1320
55	TTTATTGTTG ATTACTGTGA AAAGCATCCT GAGGCCGCC AGGAAATAGA AGTTGAGCTA	1380
	GAACCTTAGTA AAGAGATGGT TACTCTCTC CCAACAAAAA TGGAGAGATT CAGAACCAAG	1440
	ATGGCCCTGA CTTAACCTT GTTTTAAAG AAAGGAAATG TGCAATATTG AAGTGTATTT	1500
60	TTTCCCTAGT CAGACAGGCC CAATTCCATT GTGATGTTA CCTTTATAGC CAGGTGAGTG	1560

	CAGTTGAAC TTGAGATACA GTCAACTGAG TGTTTGCTAG GATCCTAAGG AACATAAAAGT	1620
5	TAATTAAAAA CTTACACCTA ATTATGTAAA TTGCTTGTT AAAGACATGT GATTGTATT	1680
	TTAGATGCTT GTTCTTATT AAAATACAGA CATTCTACC CTCAGTTCT AAATGTAGAC	1740
	TATTGTTGG CTAGTACTTG ATAGATTCT TGTAAAGAAA AATGCTGGT AATGTACCTG	1800
10	GTAACAAGCC TGTTAATATA TTAAGATTGA AAAAGTAAC TCTATAGTTA CTCCTCTAA	1860
	AATATITGAC TTCCTACATT CCCCCCACCC AAAATCTTC CCTTTGAAA ATACTAAAAA	1920
15	CTAAGTTATG TTATTATAAA GTGTAAATG GTTTGTCTTA ATTATAGGAG AAAAGGCCT	1980
	TGTTAGAAAT AAAATAAACT GACTTATTTC ACTAATGAAA AAAAAAAAAA AAAAAAAAAA	2040
	TT	2042
20		

## (2) INFORMATION FOR SEQ ID NO: 22:

25	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1872 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:	
	GGGTGACCC ACGCGTCCGA TTGGCCTAGA GCTCTGTGA CCGAGAGCGC CACGGAAAGCC	60
35	TGGGGATGAT GTGGGGCAGC TTTATTCTTT GCCTGGCTTT GGTAACTAGG TGGTCCCCTC	120
	AAGCATCTC AGTTCTCTT GCTGTTATG AATCTAAGAC AAGGAAGTCC TATAGAAGCC	180
40	AAAGGGACAG GGACGGAAAG GACAGGTCCC AAGGGATGGG GCTGTCTTTA CTTGTGGAAA	240
	CCAGGAAATT GCTCTCTCA GCCAACCAAG GTTGACCACA CACCACCTT CCGGAGCAGC	300
	TCAGTCAGCC CTCGGGACG RGAAACCACA AGCGCAGAGA CGCTGAGGCC CAGGCAGGTG	360
45	AAGAGGAAGT GGCTTTGGGT TTTTAAAGTA GGTGAGCGTG ACCTCTCTGA CTGCTCTTC	420
	CCCGGGGGGG ACTGCAAACC GCTCAGGGTT GCGGCAGAGC CATGGACTTC CGGTCCCTGC	480
50	AACGGGTGAC CTAAGCGTGG TGCAACCATC AGTCACGCAG GAGGACTGAC TTGACAGACG	540
	AAAGACAAGC CCGGATGACA CAGGGTGAGA AGAGTCAGGG CCGCACCTCT GTCCCTGCAA	600
	ACCAACAGGT GCATGGTGAG TGTGGCAGTC CCCACAGCTC CACAATGGC TCCCCCGCCA	660
55	ACGGGGACGA CAGGGATCTT CAGGAACCTTC TGACCTCACC AAGTCAAGTG GACCACTCTC	720
	CACTCCACGA GGATGTGAAA CGGTTCTTTA AAATGGGATT TTAGAGCCTC GGGAAATGCAT	780
60	GTGCGTCGCA TCTTCATAT TATGGTCAG GATAGATTCA TTTCTTGCAA CATACTGGAA	840

	AAGATATAAG CTGCAGTAAT TTGCTCTTG AATGACCGTC ACCCCCAGTA TAGGATATGC	900
	TTGTATCCCC CCGTCACTCC TCCGCCTGTT TTTTAAACTT TTCCACCACC TGCGTCCAAA	960
5	AAGAATGTTA TAGCGAGTGC TCTTAAATGT TGAACCTGGG TGTGCTTCC GGGCCAGTCT	1020
	GCGTGGCTCC ATGAAAAGCT CACTGCTGCC CCAGCCGGGC TTCTTAGAGG AGGTCAAGTTG	1080
10	TCCTATGTAT CATCATTTAC TCTGGGAATC CTACTGTGAA ATCATGTCTG TATTTTTCTG	1140
	GAGCAGTTCA CATAGAGTAG AATGTGGAAT TTCCCGTCAA CGTCTCCCTC CTCCCCCGTA	1200
	TCTGCCGCCT GTCACTTCGC CACCGTGCTA GAATACTGTT GTGTTGTAAG ATGACTAATT	1260
15	TTAAAAGAAC CTGCCCTGAA AAGTTCTTAG AAACGCAATG AAAGGGAGGA ACTTGTCCCTT	1320
	TACCCAGTTT TTCCCTTGTA GGATGGAAA GTATAAAAAG GCACAGAAGG TTGTCATGGG	1380
20	CTGTTCCCTG GGGTTTTTA TCCTGCTCAC CGTGGAGATA AGCCTGCGGC TTGTCTAACCC	1440
	AGCGCAGCGM AAAGGTCTCA ATGCCTTTG GTAACATCCG TCATTGCAGA AGAAAGTTTA	1500
	CACGACGTCA AAAAGTGACG TTCATGCTAA GTGTTTTCC AGAAATATTG GTTTCATGTT	1560
25	TCTTATTKG CTCGCCTCCT GTGCTTATAT CATCCAAAAA CTTTTAAAAA AGGTCCAGAA	1620
	TTCTATTTTA ACCTGATGTT GAGCACCTTT AAAACGTTCG TATGTGTGTT GCACTAAATTC	1680
30	TAAACTTTGG AGGCATTTTG CTGTGTGAGG CCGATGCCA CTGTAAGGT CCTAGAGTTG	1740
	CCTGTTGTC TCTGGAGATG GAATTAAACC AAATAAAGAG CTTCCACTGG AGGCTTGTAT	1800
	TGACCTTGTA ACTATATGTT AATCTCGTGT TAAAATAAAA TATAACTTGT GAAAAAAAAAA	1860
35	AAAAAAAAAC NT	1872

40 (2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 289 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

50	CATTTACCCA CCTATCAACA TGTTTGCTTT CTCTTTGTT GGTGAGAATG AGTGGCTTCT	60
	TGCTCCTAGC TAGAGCCAGT CCTTCCATAT GTGCTTTAGA TTCTTCCCTGT TTGTTCAAG	120
55	AATATTGCTC AAGCTATTCT TCCTCCCTGTT TCCTGCATCA GCATTTCCCC TCTCTACTAG	180
	ATCATCTCTG TCAGTAAATG AACATGTGTTG TGTTTCTCCT AGAAGTACTG TTTCTATATC	240
	TAGATAGTAC TCTAGCTAGA GTTAAAAAAA AAAAAAAAAA CCTNGGGGG	289

## (2) INFORMATION FOR SEQ ID NO: 24:

5	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 3533 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
	TTTTATTTAC TTCAAATTAA CTGTACTTTA CTCAAATAGA AAANGAATAA TTTCACATT 60
15	ATGAAGCTAC ACAATTCCAA AATACACATG CTGAGGCTCT TTTTAAGTCC GAATTGTCTA.
	GTAATTACAA AAAAGTGAAG AGTTTACAGA TATACAAGGA AATAAAGGCG AATTATTGCA 180
20	AAGAAAACAA GTTTAATTTC ACTTTGAATG ACAACGATTT TTCTGGAAAG CAGATACTTC
	ACTCCTTAA GTTTCCACCC AAGCCACAAT AATTTCAAAC GGTCTTGCGG ATGACCCAGC 300
	TGGTCACTCT TGTTTATGTG GGGACTGGAG GTAATGAGAG CCAAAAAAAG TGCTATAAAC 360
25	CTAATTGGC TAGAGCAAGT TCACACGACA CGACCGTGCT TTAAAAACTT GCTCTCCATT
	ATGTACTTCC TTCCATCAGG TTGGGGAAAA AAAATGGTG GGGATGGTGA GTAAACACAC 480
30	CAGTGGTTTC ATCAGAGGGG AACTCACTAC TCAGGAGGTG ACGGTGACGT GGTGCCGGTC
	CCTGAAGTAC GGGCACAAGC TCCGGAGGTT GCGGGAGCTT CCGCTGCCGC CTGGAGGGAA 540
	GCCGGAGCGA CGGGGGTCAC GGCGGCGGTC AGAGGGTAAA GGTCTTGCTC CCAGCAGCCT 660
35	CCGGGGTGGG TACGTCGCCA TCTTGGATCC GCGGGACAAG AAAATCATG CGAGGGAGAC
	GTGGTGGCG GTCCCTCCTG TGACACGACC CTTGAGTGAC AGTTCTATTT GATTGCCTCC 780
40	GGTACTGTGA GGAAAGGACA CGACTCTATG GTGAGGACTG ATGGACATAC ATTATCTGAG
	AAAAGAAACT ACCAGGTGAC AAACAGCATG TTTGGTGCTT CAAGAAAGAA GTTTGTAGAG 900
	GGGGTCGACA GTGACTACCA TGACGAAAC ATGTACTACA GCCAGTCCTC TATGTTCCA 960
45	CATCGTCAG AAAAAGATAT GCTGGCATCA CCATCTACAT CAGGTGACGT GTCTCAGTTT
	GGGGCAAGTT TATACGGCA ACAAAAGTGCCTAGG CAAATGAGGG GATGAGCAAC 1080
50	AATACCCCTC AGTTAAATCG CAGCTTATCA CAAGGCACTC AGTTACCGAG CCACGTACG
	CCAACAAACAG GGGTACCAAC AATGTCACTT CACACGCCCTC CATCTCCAAG CAGGGGTATT 1140
	TTGCCTATGA ATCCTARGAA TATGATGAAC CACTCCCAGG TTGGTCAGGG CATTGGAATT 1260
55	CCTAGCAGGA CAAATAGCAT GAGCAGTTCA GGGTTAGGTA GCCCCAACAG AAGCTCGCCA
	AGCATAATAT GTATGCCAAA GCAGCAGCCT TCTCGACAGC CTMTTACTGT GAACAGTATG 1380
60	TCTGGATTTG GAATGAACAG GAATCAGGCA TTTGGAATGA ATAACCTT ATCAAGTAAC 1440

	ATTTTTAATG GAACAGACGG AAGTGAAAAT GTGACAGGAT TGGACCTTTC AGATTTCCCA	1500
	GCATTAGCAG ACCGAAACAG GAGGGAAGGA AGTGGTAACC CAACTCCATT AATAAACCCC	1560
5	TTGGCTGGAA GAGCTCCTTA TGTTGGAATG GTAACAAAAC CAGCAAATGA ACAATCCCAG	1620
	GACTTCTCAA TACACAATGA AGATTTCCA GCATTACAG GCTCCAGCTA TAAAGATCCA	1680
10	ACATCAAGTA ATGATGACAG TAAATCTAAT TTGAATACAT CTGGCAAGAC AACTTCAAGT	1740
	ACAGATGGAC CCAAATTCCC TGGAGATAAA AGTTCAACAA CACAAAATAA TAACCAGCAG	1800
	AAAAAAAGGGA TCCAGGTGTT ACCTGATGGT CGGGITACTA ACATTCCTCA AGGGATGGTG	1860
15	ACGGACCAAT TTGGAATGAT TGGCCTGTTA ACATTTATCA GGGCAGCAGA GACAGACCCA	1920
	GGAATGGTAC ATCTTGCATT AGGAAGTGAC TTAACAAACAT TAGGCCTCAA TCTGAACTCT	1980
20	CCTGAAAATC TCTACCCCAA ATTTGCGTCA CCCTGGCAT CTTCACCTTG TCGACCTCAA	2040
	GACATAGACT TCCATGTTCC ATCTGAGTAC TTAACGAACA TTCACATTAG GGATAAGCTG	2100
	GCTGCAATAA AACTTGGCCG ATATGGTGA GACCTTCTCT TCTATCTCTA TTACATGAAT	2160
25	GGAGGAGACG TATTACAAC TTTAGCTGCA GTGGAGCTTT TTAACCGTGA TTGGAGATAC	2220
	CACAAAGAAG AACGAGTATG GATTACCAAG GCACCAGGCA TGGAGCCAAC AATGAAAACC	2280
30	AATACTATG AGAGGGGAAC ATATTACTTC TTTGACTGTC TTAACTGGAG GAAAGTAGCT	2340
	AAGGAGTTCC ATCTGGAATA TGACAAATTA GAAGAACGGC CTCACCTGCC ATCCACCTTC	2400
	AACTACAACC CTGCTCAGCA AGCCTCTAA AAAAAAAA AAAAAAAA AAAAGACTT	2460
35	CCCTTTCTT GGGTATGGC TGTCTCAGCA CAATACTCAA CATAACTGCA GAACTGATGT	2520
	GGCTCAGGCA CCCTGGTTTT AATTCCCTGA GGATCTGGCA ATTGGCTTAC GCAAAAGGTC	2580
40	ACCATTGAG GTCCTGCCCTT ACTAATTATG TGCTGCCAA CAACTAAATT TGTAATTGT	2640
	TTTTCTCTAG TTTGAGCAGG GTCTGAATT TTTCATTTAT TTCCCTTTTTT GCCAGCAGAC	2700
	AGACTTGAGT CTGTAAAGAC AAGCAAATAC ACTGACAGAA GTTTACCATA GTTTCTAAAA	2760
45	TGTAAAAAAG AAAACCCCCA AAAGACTCAA GAAAATTAGA CCACAAATTG TGCATTGTT	2820
	ATTGTAGCAC TATTGGTAAT AAAATAACAA ATGTTTGTGC ATTTTATGT GAAGATCCTT	2880
50	CTCGTATTTTC ATTTGGAAAG ATGAGCAAGA GGTCTGCTTC CTTCATTTTA CTTCCCCCTTC	2940
	TGTTTTGAA AGGCAGTTTC GCCAAGCTTA ATGCAAGAAT ATCTGACTGT TTAGAAGAAA	3000
	GATATTGCCA CAATCTCTGG ATGGTTTCC AGGGTTGTGT TATTACTGAG CTTCATCTTT	3060
55	CCAGAATGAG CAAAACACTG TCCAGTCTTT GTTACGATTT TGTAAATAAT GTGTACATTT	3120
	TTTTTAAATT TTTGGACATC ACATGAATAA AGGTATGTAT GTACGAATGT GTATATATTA	3180
60	TATATATGAC ATCTATTTG GAAAATGTTT GCCCTGCTGT ACCTCATTTC TAGGAGGTGT	3240

	GCATGGATGC AATATATGAA AATGGGACAT TCTGGAAC TG CTGGTCAGGG GACTTTGTGCG	3300
	CCCTGTGCAC TAAAAGGGCC AGATTTTCAG CAGCCAAGGA CATCCATAACC CAAGTGAATG	3360
5	TGATGGGACT TAAAAGAAGT GAACTGAGAC AATTCACTCT GGCTGTTGAC ACAGCAGCGT	3420
	TTCATAGGAA GAGAAAAAAA GATCAATCTT GTATTTCTG ACCACATAAA GGCTTCCTCT	3480
10	CTTTGTAATA AAGTAGAAAA GCTCTCCTCA AAAAAAAAAA AAAAAAAACTC GAG	3533

15 (2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1148 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

25	ACCCACCGGT CCGCAAATTA TACTTCCTCA TTTCATATTAT GTTGATACAA AAGACCTTGG	60
	CAGCCATTTTC TCCCAGCAGT TTTAAAGGAT GAACATTGGA TTTCATGCCA TCCCATAGAA	120
	AACCTGTTT AAAATTTAG GGATCTTAC TTGGTCATAC ATGAAAAGTA CACTGCTTAG	180
30	AAATTATAGA CTATTATGAT CTGTCACAG TGCCCATTGT CACTTCCTTG TCTCATTTCT	240
	TCCCTTTGTT CCTTAGTCAT CCAAATAAGC CTGAAAACCA TAAGAGATAT TACTTTATTG	300
	AATATGGTT GCATTAATT TAGCATTCA TTATCTAACAA AAATTAATAT AAATCCAGG	360
35	ACATGGTAA ATGTGTTTA ATAACCCCCA GACCCAAATG AAAATTCAA AGTCAATACC	420
	AGCAGATTCAGA TGAAAGTAA TTTAGTCCTA TAATTTTCAG CTTAATTATA AACAAAGGAA	480
40	CAAATAAGTG GAAGGGCAGC TATTACCA CGCTTAGTCAA AAACATTGGG TTACTGCCCT	540
	TTAATACACT CCTATCATCA GCACCTCCAC CATGTATTAC AAGTCTTGAC CCATCCCTGT	600
	CGTAACCTCA GTAAAAGTTA CTGTTACTAG AAAATTTTA TCAATTAACT GACAATAGT	660
45	TTCTTTTTAA AGTAGTTTCT TCCATTTTA TTCTGACTAG CTTCCAAAT GTGTTCCCTT	720
	TTTGAATCGA GGTTTTTTTG TTTTGTGTTG TTTCTGAAA AAATCATACA ACTTTGTGCT	780
50	TCTATTGCTT TTTTGTGTTT TGTTAAGCAT GTCCCTTGGC CCAAATGGAA GAGGAATGT	840
	TTAATTAAATG CTTTTAGTT TAAATAAATT GAATCATTTA TAATAATCAG TGTTAACAT	900
	TTAGTGACCC TTGGTAGGTT AAAGGTTGCA TTATTTATAC TTGAGATTTT TTTCCCTAA	960
55	CTATTCTGTT TTTTGTACTT TAAAACATAG GGGGAAATAT CACTGGCTG TCAAGAAACA	1020
	GCAGTAATTAA TTACTGAGTT AAATTGAAAA GTCCAGTGGA CCAGGCATTT CTTATATAAA	1080
60	TAAAATTGGT GGTACTAATG TGAAAAAAA AAAAAAAAAA AACTCGAGGG GGGCCCGGTA	1140

CCCTATTA

1148

5

## (2) INFORMATION FOR SEQ ID NO: 26:

- 10                   (i) SEQUENCE CHARACTERISTICS:  
                       (A) LENGTH: 717 base pairs  
                       (B) TYPE: nucleic acid  
                       (C) STRANDEDNESS: double  
                       (D) TOPOLOGY: linear

## 15                   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GGCACGAGCT AGCTGCCGCC ACCCGAACAG CCTGTCTGG TGCCCCGGCT CCCTGCCCG	60
CGCCCAGTCA TGACCCCTGCG CCCCTCACTC CTCCCGCTCC ATCTGCTGCT GCTGCTGCTG	120
CTCAGTGCAG CGGTGTGCCG GGCTGAGGCT GGGCTCGAAA CCGAAAGTCC CGTCCGGACC	180
CTCCAAGTGG AGACCCCTGGT GGAGCCCCCA GAACCATGTG CCGAGCCCGC TGCTTTGGA	240
25                 GACACGCTTC ACATAACACTA CACGGGAAGC TTGGTAGATG GACGTATTAT TGACACCTCC	300
CTGACCAGAG ACCCTCTGGT TATAGAACATT GGCCAAAAGC AGGTGATTCC AGGTCTGGAG	360
30                 CAGAGTCTTC TCGACATGTG TGTGGAGAG AAGCGAAGGG CAATCATTCC TTCTCACTTG	420
GCCTATGGAA AACGGGGATT TCCACCATCT GTCCCGAGGG ATGCAGTGGT GCAGTATGAC	480
GTGGAGCTGA TTGCACTAAT CCGAGCCAAC TACTGGCTAA AGCTGGTGAA GGGCATTTTG	540
35                 CCTCTGGTAG GGATGGCCAT GGTGCCAGCC CTCTGGGCC TCATTGGTA TCACCTATAC	600
AGAAAGGCCA ATAGACCCAA AGTCTCCAAA AAGAAGCTCA AGGAAGAGAA ACGAAACAAG	660
40                 AGCAAAAAGA AATAATAAAT AATAAATTAA AAAAAAAA AAAAAAAA AAAAAAA	717

## 45                   (2) INFORMATION FOR SEQ ID NO: 27:

- 45                   (i) SEQUENCE CHARACTERISTICS:  
                       (A) LENGTH: 1099 base pairs  
                       (B) TYPE: nucleic acid  
                       (C) STRANDEDNESS: double  
                       (D) TOPOLOGY: linear

## 50                   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

55                 GGCACGAGCC GATGTGGACA TCATCCTGTC TATCCCCATG TTCCTGCCGCC TGTACCTGAT	60
CGCCCGAGTC ATGCTGCTGC ACAGCAAGCT CTTCACCGAT GCCTCGTCCC GCAGCATCGG	120
GGCCCTCAAC AAGATCAACT TCAACACCCG CTTTGTATG AAGACGCTCA TGACCATCTG	180
60                 CCCTGGCACT GTGCTGCTCG TGTTCAGCAT CTCTCTGTGG ATCATTGCTG CCTGGACCGT	240

	CCGTGTCGT GAAAGTCCTG AATCACCAAGC CCAGCCTTCT GGCTCATCAC TTCCCTGCTTG	300
5	GTACCATGAC CAGCAGGACG TAACTAGTAA CTTTCTGGGT GCCATGTGGC TCATCTCCAT	360
	CACATTCCCTT TCCATTGGTT ATGGGGACAT GGTGCCCCAC ACATACTGTG GGAAAGGTGT	420
	CTGTCTCCTC ACTGGCATCA TGGGTGCAGG CTGCACTGCC CTTGTGGTGG CCGTGGTGGC	480
10	CCGAAAGCTG GAACTCACCA AAGCGGAGAA GCACGTTCAT AACTTCATGA TGGACACTCA	540
	GCTCACCAAG CGGATCAAGA ATGCTGCAGC CAATGTCCCTT CGGGAAACAT GGTTAACATCA	600
15	TAAACACACA AAGCTGCTAA AGAAGATTGA CCATGCCAAA GTGAGGAAAC ACCAGAGGAA	660
	GTTCCTCCCA AGCTATCCAC CAGTTTGAGG AGCGTCCCAG ATGGAACAGA GGAAAGCTGA	720
	GTGACCAAGC CAACACTCTG GTGGACCTTT CCAAGATGCA GAATGTCAATG TATGACTTAA	780
20	TCACAGAACT CAATGACCGG AGCGAAGACC TGGAGAAGCA GATTGGCAGC CTGGAGTCGA	840
	AGCTGGAGCA TCTCACCGCC AGCTTCAACT CCCTGCCGCT GCTCATGCC GACACCCCTGC	900
25	GCCAGCAGCA GCACCGAGCTC CTGTCTGCCA TCATCGAGGC CCGGGGTGTC AGCGTGGCAG	960
	TGGGCACCAC CCACACCCCCA ATCTCCGATA GCCCCATTGG GGTCAGCTCC ACCTCCCTCC	1020
	CGACCCCGTN CACAAGTTCA AGCAGTGTGCT AAATAAATCT CCCCACTCCA GAAGCAITAA	1080
30	AAAAAAAAA AAAAAAAA	1099

35 (2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
- (B) TYPE: nucleic acid
- 40 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

45	AATTCGGCAG AGAGCCAACC GAGGGCGTTC CTGTCGGGGC TGCAAGCGGCG GGAGGGAGCC	60
	CAGTGGAGGC GCCCTCCCGA ACCGCCACTG CCCATGCTGA CCACCCAGCC CTCCGGCTGC	120
50	TGATGTCAATG AGTAACACCA CTGTGCCCAA TGCCCCCCAG GCCAACAGCG ACTCCATGGT	180
	GGGCTATGTG TTGGGGCCCT TCTTCCTCAT CACCCGGTC GGGGTGGTGG TGGCTGTGGT	240
	AATGTATGTA CAGAAGAAAA AGCGGGGTGGA CCGGCTGCGC CATCACCTGC TCCCCATGTA	300
55	CAGCTATGAC CCAGCTGAGG AACTGCATGA GGCTGAGCAG GAGCTGCTCT CTGACATGGG	360
	AGACCCCAAG GTGGTACATG GCTGGCAGAG TGGCTACCAAG CACAAGCGGA TGCCACTGCT	420
60	GGATGTCAAG ACGTGACCTG ACCCCCTTGC CCCACCCCTTC AGAGCCTGGG GTYCTGGACT	480

	GCCTGGGCC CTGCCATCTG CTTCCCTGC TGTCACCTGG STCCCCITGC TGGGTGCTGG	540
	GTCTCCATTG CTCCCTCCAC CCACCCCTCAG CAGCATCTGC TTCCCATGCC CTCACCACCA	600
5	CCTCACTGCC CCCAGGCCTT CTGCCCTTIG TGGGTGTTGA GCTCACCGCC CACCCACAGG	660
	CACTCATGGG AAGAGGCTTT CCTTCTGGGA TGGCGGCGGC TGGTAGACAC CTTTGCTTTC	720
10	TCTAGCCCTC CTGGCTGGG CTTGGGCACA AATCCCAGG CAGGCTTGG AGTTGTTCC	780
	ATGGTGATGG GGCCAGATGT ATAGTATTCA GTATATATTT TGTAAATAAA ATGTTTGTG	840
	GCTAAAAAAA AAAAAAAAAA ATCNAAGGGG GGGCCGGTAC CCAAATTCCC CCTATANTGA	900
15	ATTCGTATTA ACAATTCACT TGGGGCCGTC CTTTTAANAA C	941

## 20 (2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 756 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

30	GGCACGAGGA AGCTGGAGCG GGCGGGCGGT GCAGTCACGG GGGAGCGAGG CCTGCTGGC	60
	TTGGCAACGA GGGACTCGGC CTCGGAGGCG ACCCAGACCA CACAGACACT GGGTCAAGGA	120
35	GTAAGCAGAG GATAAACAAAC TGGAAGGAGA GCAAGCACAA AGTCATCATG GCTTCAGCGT	180
	CTGCTCGTGG AAACCAAGAT AAAGATGCCA ATTTTCCACC ACCAAGCAAG CAGACCTGT	240
	TGTTTGTCC AAAATCAAAA CTGCACATCC ACAGAGCAGA GATCTCAAAG ATTATGCGAG	300
40	AATGTCAAGGA AGAAAGTTTC TGGAAGAGAG CTCTGCCTTT TTCTCTGTGTA AGCATGCTTG	360
	TCACCCAGGG ACTAGTCTAC CAAGGTTATT TGGCAGCTAA TTCTAGATTT GGATCATTGC	420
45	CCAAAGTTGC ACTTGCTGGT CTCTGGGAT TTGGCCTTGG AAAGGTATCA TACATAGGAG	480
	TATGCCAGAG TAAATTCCAT TTTTTTGAGA ATCAGCTCCG TGGGGCTGGT TTTGGTCCAC	540
	AGCATAACAG GCACTGCCTC CTTACCTGTG AGGAATGCCA AATAAACAT GGATTAAGTG	600
50	AGAAGGGAGA CTCTCAGCCT TCAGCTTCCT AAATTCGTG TCTGTGACTT TCGAAGTTTT	660
	TTAAACCTCT GAATTGTAC ACATTTAAAA TTTCAAGTGT ACTTTAAAT AAAATACTTC	720
55	TAATGGAAAA AAAAAAAAAA AAAAAAAAAA ACTCGA	756

## 60 (2) INFORMATION FOR SEQ ID NO: 30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2100 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 5 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

10	NCCAGAGGCA GAAAGTCCTG CTTCTGGGGC GTAACCTACA GGATATCCTT GGAACAGAAC	60
	ATCTTATTGT GGAAAGTRACT TCCAATGATG CTGTGAGATT TTATCCCTGG ACCATTGATA	120
	ATAAAATACTA TTCAGCAGAC ATCAATCTAT GTGTGGTGCC AAACAAATTG CTTGTTACTG	180
15	CAGAGATTGC AGAACATCTGTC CAAGCATTTG TGGTTTACTT TGACAGCACA CAAAATCGG	240
	GCCTTGATAG TGTCTCCTCA TGGCTTCCAC TGGCAAAAGC ATGGTTACCCY GAGGTGATGA	300
20	TCTTGGTCTG CGATAGAGTG TCTGAAGATG GTATAAACCG ACAAAAAGCT CAAGAATGGT	360
	GCATCCAAAC ATGGCTTTGA ATTGGTAGAA CTTAGTCCAG AGGAGTTGCC TGAGGAGGAT	420
	GATGACTTCC CAGAATCTAC AGGAGTAAAG CGAATTGTCC AAGCCCTGAA TGCCAATGTG	480
25	TGGTCCAATG TAGTGATGAA GAATGATAGG AACCAAGGCT TTAGCTTGCT GCAACTCATT	540
	GAATGGAAACA AACCATAGCA TTGGGTCAAGC AGATCCCTGT CACCCAGAGC AACCCCATTT	600
30	GCCAGCAGCA GATAGTACTG AATCCCTCTC TGATCATCGG GGTGGTGCAT CTAACACAAAC	660
	AGATGCCAG GTTGATAGCA TTGTGGATCC CATGTTAGAT CTGGATATTG AAGAATTAGC	720
	CAGTCCTTACC ACTGGAGGAG GAGATGTGGA GAATTTGAA AGACTCTTT CAAAGTTAAA	780
35	GGAAATGAAA GACAAGGCTG CGACGCTTCC TCATGAGCAA AGAAAAGTGC ATGCAGAAAA	840
	GGTGGCCAAA GCATTCTGGA TGGCAATCGG GGGAGACAGA GATGAAATTG AAGGCCTTTC	900
40	ATCTGATGAA GAGCACTGAA TTATTCATAC TAGGTTTGA CCAACAAAGA TGCTAGCTGT	960
	CTCTGAGATA CCTCTCTACT CAGCCCAGTC ATATTTGCC AAAATTGCC TTATCATGTT	1020
	GGCTGCCTGA CTTGTTTATA GGGTCCCCTT AATTTAGTT TTTAGTAGGA GGTTAAGGAG	1080
45	AAATCTTTTT TTTCTCTAGT ATATTTGAAG AGAGTGAGGA ATACAGTGAT AGTAATGAGT	1140
	GAGGATTTCT TAAATRTACT TTTTTTTGT TCTAGGAATG AGGGTAGGAT AAATCTCAGA	1200
50	GGTCTGTGTG ATTTACTCAA GTTGAAGACA ACCTCCAGGC CATTCTGGT CAACCTTTA	1260
	AGTAGCATTT CCAGCATTCA CACTTGATAC TGCACATCG GAGTTGTGTC ACCTTTCTG	1320
	GGTGATTTGG GTTTCTCCA TTCAAGGAGC TTGTAGCTCT GAAGCTATGA TGCTTTTATT	1380
55	GGGAGGAAAG GAGGCAGCTG CAGAATTGAT GTGAGCTATG TGGGGCCGAA GTCTCAGCCC	1440
	GCAGCTAAGT CTCTACCTAA GAAAATGCCT CTGGGCATTC TTTTGAAGTA TAGTGTCTGA	1500
60	GCTCATGCTA GAAAGAATCA AAAAGCCAGT GTGGATTTTT AGACTGTAAT AAATGAGGCA	1560

	AAGGATTCT ATTCCAGTGG GAAGRAAAACC TCTCTACTGA GTTGTGGGG ATATGTGT	1620
	TGTTAGAGAG AACCTTAAGG AGTCCTTGTA TGGGCCATGG AGACAGTATG TGATAACATA	1680
5	CCGTGATTT CATGAAGAAA TTCTTCTGTC TTAGAGTTCT CCCCTGCTGC TTGAGATGCC	1740
	AGAGCTGTGT TGTTGCACAC CTGAAAACA AGGCACATTT CCCCCCTTCT CTTAAAGCC	1800
10	AAAGAGAGAT CACTGCCAAA GTGGGAGCAC TAAGGGGTGG GTGGGAAAGT GAAATGTTAG	1860
	GCGATGAATT CCTGAGCACC TTGTTTTCTC TCCAAGGTC GTAGCTCCTC TCTGCCCTTC	1920
	CAAGCCTGTA ACCTCGGAGG ACTATCTTTT GTTCTTTATC CTTTGTCTTG TTTGAGTGGG	1980
15	TCAGCCCCAG AGGAACGTAT AAGCAAATGG CAAGTTTTA AAGGAAGAGT GGAAAGTACT	2040
	GCAAATAAAA ATCCTTATTT GTTTTGTAG AAAAAAAAAA AAAAAAAAAA AAAAAAAAAAG	2100

20

## (2) INFORMATION FOR SEQ ID NO: 31:

## (i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 1448 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

	AAAAAAAAAA AAAGCCCACC TGAAAGCCTG TCTCTTTCCA CTTTGTGGC CCTTCCAGTG	60
	GGATTATCGA GCATGTTGTT TTTCATAGT GCCTTTTTCC TTATTTCAAG GGTTGCTTCT	120
35	GAGTGGTGT TTTTTTTTTT TTAATTGTT TTGTTTTAAA ATAAGTTAAA GACAGTCCAG	180
	AGCTTTTCAG CCAATTGTC TCCTACTCTG TGTAAATATT TTTCCCTCCG GGCAGGGAG	240
40	CCAGGGTAGA GCAAAGGAGA CAAGCAGGAG TGGAAAGGTGA GGCGTTCTCC TGCTTGACT	300
	AAGCCAGGAG STTTAAGCTC CAGCTTTAAC GGTTGTGAGC CCCTGGGGT TCAGGAACT	360
	GCTTGCCAG GGTGCACTGT GAGTGTGATG GGCCACCGGG GCAAGAGGGA AGGTGACCGC	420
45	CCAGCTCTCC CACATCCAC TGGATCTGGC TTACAGGGGG GTCGGAAGCC TGTCTCACC	480
	GTCTGGGGG TTGTGGCCCC CGCCCCCTCC CTATATGCAC CCCTGGAACCC AGCAAGTCCC	540
50	AGACAAGGAG AGCGGAGGAG GAAGTCATGG GAACGCAGCC TCCAGTTGTA GCAGGTTCA	600
	CTATTCCAT GCTGGGGTAC ACAGTGTGAGAG TACTCACTTT TCACTTGCTC TGCTCTTACA	660
	TTGGGCCATG GCTTTCATCC TGTGTCCCCCT GACCTGTCCA GGTGAGTGTG AGGGCAGCAC	720
55	TGGGAAGCTG GACTGCTGCT TGTGCCTCCC TTCCCCAGTGG GCTGTGTGTA CTGCTGCTCC	780
	CCACCCCTAC CGATGGTCCC AGGAAGCAGG GAGAGTTGGG GAAGGCAAGA TTGGAAAGAC	840
60	AGGAAGACCA AGGCCTCGGC AGAACTCTCTC GTCTTCTCTC CACTCTGGT CCCCTGTGGT	900

	GATGTGCCTG TAATCTTTT CTCCACCAA ACCCCTCCC ACGACAAAAA CAAGACTGCC	960
5	TCCCTCTCTT CCGGGAGCTG GTGACAGCCT TGGCCTTTC AGTCCCAAAG CGGCCGATGG	1020
	GAGTCTCCCT CCGACTCCAG ATATGAACAG GGCCCAGGCC TGGAGCGTTT GCTGTGCCAG	1080
	GAGGCAGGCAG CTCTCTGGG CAGAGCCTGT CCCCGCTTC CCTCACTCTT CCTCATCCTG	1140
10	CTTCTCTTTT CCTCGCAGAT GATAAAAGGA ATCTGGCATT CTACACCTGG ACCATTTGAT	1200
	TGTTTTATTT TGGAATTGGT GTATATCATG AAGCCTTGCT GAACTAAGTT TTGTGTGTAT	1260
15	ATATTTAAAAA AAAAAATCAG TGTTAAATA AAGACCTATG TACTTAATCC TTTAACTCTG	1320
	CGGATAGCAT TTGGTAGGTA GTGATTAATC GTGAATAATA AATACACAAT GAATTCTIMA	1380
	AAAAAAAAAAA AAAAAAAAAAA AAAAAAAAAAA AAACCCCGGG GGGGGCCCCG GGCCCCAATT	1440
20	CCCCCCAA	1448

## 25 (2) INFORMATION FOR SEQ ID NO: 32:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 456 base pairs
- (B) TYPE: nucleic acid
- 30 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

35	GGCACAGCAA ACTTGACGCC ATGAAGATCC CGGTCCCTCC TGCCGTGGTG CTCCCTCTCCC	60
	TCCTGGTGCT CCACTCTGCC CAGGGAGCCA CCCTGGGTGG TCCTGAGGAA GAAAGCACCA	120
40	TTGAGAAATTA TCGGTACAGA CCCGAGGCCT TTAACACCCC GTTCCTGAAC ATCGACAAAT	180
	TGCGATCTGC GTTTAAGGCT GATGAGITCC TGAACGGCA CGCCCTCTTT GAGTCTATCA	240
	AAAGGAAACT TCCTTCCTC AACTGGGATG CCTTCCTAA GCTGAAAGGA CTGAGGAGCG	300
45	CAACTCCTGA TGCCCAGTGA CCATGACCTC CACTGGAAGA GGGGGCTAGC GTGAGCGCTG	360
	ATTCTCAACC TACCATAACT CTTCTCTGCC TCAGGAACTC CAATAAAACA TTTTCCATCC	420
50	AAAAAAAAAAA AAAAAAAAAAC CCCNGGGGGG GCCCGG	456

## 55 (2) INFORMATION FOR SEQ ID NO: 33:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1326 base pairs
- (B) TYPE: nucleic acid
- 60 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

	GGCACCGACTG CAGGCCAGA GAGGACTCAT TGAAAGGACT GAAAGGGGAG GTGGCGTTTT	60
5	CTTCCTACCC AAACCTTACCC CTGTGAGCTG GACAGCTTGG TAGCACCTGC CTGGACTTAG	120
	ATGGTGGTAG CCAAGAACAC TGACATTITA GGGAACAGGA CGGGGAGGAG AAGGCTCTGG	180
10	CACACACACA TGTGTCCATA TGTCTGCAA TGGTCTGGGG ACTATTGCTA GGCTAGGAGC	240
	CCTAAGTGTC TTCTTCCCTCA TGTCTTTCT CCCCTGTSTC ATGGGCCCTA AGRTCTCTTT	300
15	CACTGGGCCT GCCTCAATGA ACGTGCTGCC CAGCTACCCC GAAACACGGC ANCTGCCGGC	360
	TATCAATGCC CCAGCTGCAA TGGCCCATCT TCCCCAACC AACCTGGCTG GGCCCGTGGG	420
	CTCCGCACTG AGARARAAAAS TTGGCACART CAACTGGGCC CGGGCAGGAC TGGGCCYCCC	480
20	TCTGATCGAT GAAGKTCGTG ARCCCAGAGC CCGAGCCCCCT CAACACGTCT GACTTCTCTG	540
	ACTGGTCTAG TTTTAATGCC AGCAGTACCC CTGGACCAGA CGAGGGTAGAC AGCGCCTCTG	600
25	CTGGCCCCAGC CTTCTACAGC CGAGCCCCCC GGCCCCCAGC TTCCCCAGGC CGGCCCGAGC	660
	AGCACACAGT GATCCACATG GGCAATCCTG AGCCCTTGAC TCACGCCCT AGGAAGGTGT	720
	ATGATACCGG GGATGATGAC CGGACACCAAG GCCTCCATGG AGACTGTGAC GATGACAAGT	780
30	ACCGACGTCG GCCGGCCTTG GGTTGGCTGG CCCGGCTGCT AAGGAGCCGG GCTGGGTCTC	840
	GGAAGGGRCC GCTGACCCCTG CTCCACGGGG CGGGGCTGCT GCTACTCTTG GGACTGCTGG	900
	GCTTCCTGGC CCTCCTTGCC CTCATGTCTC GCCTAGGCCG GGCGCAGCT GACAGCGATC	960
35	CCAACCTGGA CCCACTCATG AACCCCTCACA TCCCGCTGGG CCCCTCCTGA GCCCCCTTGC	1020
	TTGTGGCTAG GCCAGCCTAG GATGTGGTT CTGTGGAGGA GAGGGGGGT AATGGGGAGG	1080
40	CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC	1140
	AAAGCCAAGT CCACCAAGAGT GGCTGCAGGC CAGGCCTGGA GTCCCCGTGG GTCAAGCAAT	1200
45	TGTCTTGTACT TGCTTTCTC CGGGGTYTCC AGCCTCCGAC CCCTCGCCCC ATGAAGGAGC	1260
	TGGCAGGTGG AAATAAACAA CAACTTTATT AAAAAAAA AAAAAAAA AAAAAAAA	1320
	AAANAA	1326
50		

## (2) INFORMATION FOR SEQ ID NO: 34:

55       (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 710 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

	GCGAAAGAGA AAAAGGCTGG AGCTCCGCC CCCGGGGCTG TCAGATGGCT TGGGTTCCTG	60
5	CGACCGGATT GGCTCGCGGA GGGCAGAAAT TACTCAGCAA ACATGACTAT TATTAGCTGC	120
	TTAGAACAG CTCACCAAAG TAGAGAGACC ACCCAGGTAG GCAACCCAGT GTGTGCATCC	180
10	TCGGCTTCGG GGCAGCCTCT GAGAGCGCCA ACCTTCCTCGC ATGCAATACT TCCATTAAGG	240
	AATGCTCCCC CTCCTTCTC TCTTATTCTCT TTTCTTTCA ACAGTGTCTT CTTTTGTGG	300
	GATGCCTTTG CGCGCACACA CGCGCGCGCA SGCACACACA CGAACATTG CCTCGGGTA	360
15	GACACGGGGG GAAATGTWAT ATTTTTTTAA GCGCTTAAAC AATTTCTGAA ATTCTCTAAA	420
	GAAAAGCCTT TCAGARGCAC CTTGGCCTCA AGCTGCAACA AATACTGGGA RGTCCGGCTC	480
20	GCATTCCCAG GCCTGCACCA ATAATGACAG CGTGCTGGAT ARTGCGCCAG TGTGTGCCAG	540
	ATTTTTTTTT CCTCTTCTCT TTTCCTTTAT AACTAAAGGG AAGACTTAGG CTCTTGCAGG	600
	GAACAACGCC TCGCATTAAG ATAAACAGAA TGGAAAGTTA AAGAGGAAAG CAAGGACGTT	660
25	GGGAAAAGCC ATCTTCTTA AAATCCGTCT GCCCCCCAGC CGCTTTCTCC	710

## 30 (2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1188 base pairs
- (B) TYPE: nucleic acid
- 35 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

40	GATGGCTTTT ATATCTATTA TCGACCCACA GACAGTGACA ATGATAGTGA CTACAAGAAG	60
	GATATGGTGG AAGGGGACAA GTACTGGCAC TCCATCAGCC ACCTGCAGCC AGAGACCTCC	120
45	TACGACATTA AGATGCAGTG CTTCATGAA GGAGGGGAGA GCGAGTTCAG CAACGTGATG	180
	ATCTGTGAGA CCAAAGCTCG GAAGTCTTCT GGCCAGCCTG GTCGACTGCC ACCCCCAACT	240
	CTGGCCCCAC CACAGCCGCC CCTTCCTGAA ACCATAGAGC GGCCGGTGGG CACTGGGCC	300
50	ATGGTGGCTC GCTCCAGCGA CCTGCCCTAT CTGATTTGTCG GGGTCGTCTT GGGCTCCATC	360
	GTTCCTCATCA TCGTCACCTT CATCCCCCTTC TGCTTGTGGA GGGCTGGTC TAAGCAAAAA	420
55	CATACAACAG ACCTGGGTTT TCCTCGAAGT GCCCTTCCAC CCTCCTGCC GTATACTATG	480
	GTGCCATTGG GAGGACTCCC AGGCCACCAAG GCAGTGGACA GCCCTACCTC AGTGGCATCA	540
	GTGGACGGGC CTGTGCTAAT GGGATCCACA TGAATAGGGG CTGCCCCCTCG GCTGCAGTGG	600
60	GCTACCCGGG CATGAAGCCC CAGCAGCACT GCCCAGGCGA GCTTCAGCAG CAGAGTGACA	660

	CCAGCAGCCT GCTGAGGCAG ACCCATTTG GCAATGGATA TGACCCCCAA AGTCACCAGA	720
5	TCACGAGGGG TCCCAAGTCT AGCCCGACG AGGGCTCTTT CTTATACACA CTGCCCCACG	780
	ACTCCACTCA CCAGCTGCTG CAGCCCCATC ACGACTGCTG CCAACGCCAG GAGCAGCCTG	840
	CTGSTGTGGG CCAGTCAGGG GTGAGGAGAG CCCCGACAG TCCTGTCCCTG GAAGCAGTGT	900
10	GGGACCCCTCC ATTTCACTCA GGCCCCCAT GCTGCTTGGG CCTTGTGCCA GTTGAAGAGG	960
	TGGACAGTCC TGACTCCTGC CAAGTGAAGTG GAGGAGACTG GTGCCCCAG CACCCCGTAG	1020
15	GGGCCTACGT AGGACAGGAA CCTGGAATGC AGCTCTCCCC GGGGCCACTG GTGCGTGTGT	1080
	CTTTTGAAAC ACCACCTCTC ACAATTAGG CAGAAGCTGA TATCCCAGAA AGACTATATA	1140
	TTGTTTTTTT TTAAAAAAA AAAAAAAA AWCYCGGGGG GGGGCCCC	1188
20		

## (2) INFORMATION FOR SEQ ID NO: 36:

25       (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 956 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

30       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

	GGCAGAGCAG TGAAAATGCA TCCTAAAAAT TCAATGTITA TACCAAGGCTC ATGACACTAA	60
35	GATGTGACAT CTGGACACGA GGGTCAGCC ACGTGGATAC ATCCCTCCCA GATTGCATCT	120
	CCAGGAATCA CTCTGCTAGC AGAATGGCG CCCCATCCCT TACTATGCTG CTCCCTCCCA	180
40	AAGTGCAGCC CAGAAGGACC CAGGCCCTTG ATGCACATTG GGTGGGTCTC CCACTACTTT	240
	AGTTGAAATG GGAGCATGCT GGAGTCGGCG TTCTGTTGCT TCTGGTGAGA AGGACATCCC	300
	ATTGACCCCT GGCCACCAAGG TCCAGTATTTC CATCCTCCCT TCTGTCCCA CCTATCGCCC	360
45	TCCCCACYAG GCCCACCCCC ACAACTCTC CTCAAGGGAG GTNTCCCGC AGCTGGAGGG	420
	CTTGCACAGA CCAGCAGTCA CAGAAATCAT TCTTCCTGCT GTACTGGGCC TTAACTGCCT	480
50	GCAAATGTCC GAGCACTACT GCATAGGATG CCAGAGCCAC CGAAGATAAA CACAGCCAAG	540
	TTTAATAATA ATAAAAGGAA AAATCTCAGC CTGCAGAACT CTGGTTTGA CCCACCATCG	600
	GCCAGATGCA CATCTTCAGG GCCTGTTGAG CACCTCTGA AAAGCAGGGC TCGTAATAGA	660
55	CTCCAGCACA TTCCATCAGA GTCAGGAAAA CTGCGGTGAG TCCCAGAGAA TCTAGGGTGC	720
	AGGGCAGGGA GCAGGAGTCA TAAGGAGTGA TAACCTAAC TGTGTGTAGT CAGCGGGGAG	780
60	GGTCTTATGT TATCAGGTGA AATGAGAGCC AGTAAGTTAG TTGATCCTGT CACAGATATA	840

ACCCCTGATAA CACCCCATAG ATACGCGACA CGTGTGCCT GCCCCTGCTT TCCCCATCCA	900
ACATGGTTCT TCTGTTCCAC AGACATTAAA GGGGCTTTCT GCAATTACTT AAAAAAA	956

5

## (2) INFORMATION FOR SEQ ID NO: 37:

- 10           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 1603 base pairs  
               (B) TYPE: nucleic acid  
               (C) STRANDEDNESS: double  
               (D) TOPOLOGY: linear

15           (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TCGACCCACG CGTCCGCTCT GCCAGGAATC TGGTCTTCT GTAGACCCAA GTCAGAAAGA	60
20 ACCATTTGTG GAGTTAACAT GAATATTAGA RGCATTAAR GTCAGAGTTC TGAGACCTGC	120
TCTGGAATGG GCAGTTCAA ACCGAGAGAT GCTTATAGCC CAAAACAGCT CCTTGAAATT	180
25 TAAACTACAC AGACTGTATT TTATTAGCTT RTTAATGGGT GGAACACAAA TCAGCGAGAR	240
GCATTACAAT ATGCTAAAAA TTTTCAGCCA TTTGCCCTAA ATCATCAAA AGACATTTCAG	300
GTTTTGATGG GAAGCCTTGT GTACCTGAGA CAAGGGATTG AGAACTCACC ATATGTTCAC	360
30 CTACTTGATG CAAACCAAGTG GGCTGATATC TGTGACATCT TTACACGGGA TGCTTGTGCC	420
CTCCTGGGC TCTCCGTGGA GTCCCCCTCTC AGTGTCACTT TCTCAGCAGG TTGTGTGGCG	480
35 CTGCCAGCTT TAATTAACAT CAAAGCCGTG ATTGAACAGA GGCAGTGTAC TGGAGTTTGG	540
AACCAGAAAG ATGAATTACC TATTGAAGTG GACCTTGGTA AAAAGTGTG GTATCACTCT	600
ATATTTGCCT GCCCATTCTC TCGTCAGCAA ACAACAGATA ACAATCCACC CATGAAATTG	660
40 GTCTGTGGTC ATATTATATC AAGAGATGCC CTGAATAAA TGTTTAATGG TAGCAAATTA	720
AAATGTCCCT ACTGTCCAAT GGAACAAAGT CCAGGAGATG CCAAACAGAT ATTTTCTGA	780
45 AGAGATAACT TTAGTTTGCAT ATTGTAAAGT GAAACTGAAT CGTGGTGCAT TTTCAGAAGA	840
GAACGTTCCA TATAATGCAG CTAACCAAGG ACTCCTGTGT TTCTATAAGC TAATGCTCCA	900
GAAACTTTGC CAACCTGTGA GTGTACACAC ACTGAGGGGA GTGCTCCCGG TGAATATTAT	960
50 CATAGGGCTT TATTATATTC TTGGTCTTCA TTTCTGATCA AGTAAATACA CCAGCAGTTG	1020
TCATTCAATG CAGTTTTTG TACTTAATTA TATGTTGATT TTTTTACTTT TTAAGAGCAG	1080
55 AAACGGAAAT TGACCTCCCC GCCATGTGTT TAATATTCCCT CCTGCTTTA CTTTTGTCAT	1140
TTTCTTGATA ATCGTAAGCC TTGAGAGTGT TTGTGAAAAA GTTTTATTC CTGTTATGTA	1200
TACATAATTA AATGAAAATT CTTCAGAAAA AGTTTGATAA ATTGAATTGT GGTTATGAAA	1260
60 CTAATTGCA TTTTTATTTG CTTAAGAAAG AAAGCTGTGA TAGATTCCAG ATATGCTTT	1320

	TGATGTTTTC CTCTGCTCCA GCTCCAAGAA GTCAGCACAC CTGCATTTA GCTCTGCATG	1380
5	CAGCCCCAGC AGGCTGCGTG TTTAAGAATT TCATGTTTA ACTGGCTGGT GTGAGAAGTC	1440
	TTCCGTTAGC ATAGAGTGGA AGGAGTACTA TTGTTTGGTT GGTTTTGTG TTGTTTGTGTT	1500
	TTTGTGTTTG CTTTATTGCA CAAGAGGTGC TTGTTTAAA AGTATGTTA ATAAAATGAA	1560
10	ATTCTAAAGT TAARAAGTGT TCCTAAAGTT GATATTTAAC TCT	1603

## 15 (2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1089 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

25	GGCACGAGCT ACCTTTCTGC CTGCTTGCT GGCTGCAACA GCACGAATCT CACGGGCTGT	60
	GCGTGCCTCA CCACCGTCCC TGCTGAGAAC GCAACCGTGG TTCCTGGAAA ATGCCCGAGT	120
30	CCTGGGTGCC AAGAGGCCTT CCTCACTTTTC CTCTGTGTGA TGTGTATCTG CAGCCTGATC	180
	GGTGCCATGG CAAGACACCC TCAGTCATCA TCCTCATCAG GACAGTCAGC CCTGAACCTCA	240
	AGTCCTTACGC TTTGGGAGTT CTTTTCTCC TCCTTCGTTT GTTGGGCTTC ATCCCTCCAC	300
35	CCCTCATCTT CGGGGCTGGC ATCGACTCCA CCTGCCTGTT CTGGAGCAGC TTCTGTGGGG	360
	AGCAAGGGCGC CTGCGTCCTC TACGACAATG TGGCTTACCG ATACCTGTAT GTCAGCATCG	420
40	CCATCGCGCT CAAATCCTTC GCCTTCATCC TGTACACCAC CACGTGGCAG TGCTGAGGAA	480
	AAACTATAAA CGCTACATCA AAAACCACGA GGGGGGGCTG AGCACCAGTG AGTTCTTTGC	540
	CTCTACTCTG ACCCTAGACA ACCTGGGGAG GGACCCCTGTG CCCGCAAACC AGACACATAG	600
45	GACAAAGTTT ATCTATAACC TGGAAGACCA TGAGTGGTGT GAAAACATGG AGTOCGTTTT	660
	ATAGTGACTA AAGGAGGGCT GAACTCTGTA TTAGTAATCC AAGGGTCATT TTTTCTTAA	720
50	AAAAAGAAAA AAAGGTTCCA AAAAAAACCA AAACTCAGTA CACACACACA GGCACAGATG	780
	CACACACACG CAGACAGACA CACCGACTTT GTCCCTTTTC TCAGCATCAG AGCCAGACAG	840
	GATTCAAAAT AAGGAGAGAA TGACATCGTG CGGCAGGGTC CTGGAGGCCA CTCGCGCGGC	900
55	TGGGCCACAG AGTCTACTTT GAAGGCACCT CATGGTTTTC AGGATGCTGA CAGCTGCAAG	960
	CAACAGGCAC TGCCAAATTC AGGAAACAGT GGTGGCCAGC TTGGAGGATG GACATTCTG	1020
60	GATACACATA CACATACAAA ACAGAAAACA TTTTTAAAAA GAAGTTCCCT AAAATAAAA	1080

AAAAAAA

1089

5

## (2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 629 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

15	AGCTCAGTTC CCTTAGAAAT GAAATTAAAT ATGACACTAC CAGGTAAGCC ACTGAGACCA	60
	GTGGAGGTGA TAGCTAAGAA CATAAGGAAT TAAGAATTAAAT GGAGGAGAA AGGAGGTAAT	120
20	GAATACCACT TACATCCTAA GACTCACTGT AGTGGTGAGT GTGTAATTT ATCTCGCTAT	180
	CCATCCTCTT TTAAGTTTTT CCTTAGAAAG TCCTCTATTG GTACCTTGGA GGGACTGCTG	240
25	TCAAAATATA TGGAAAAGTG GGTCTGTGTG GTACAAGAGG TGGACTTGC CACACATGGA	300
	AGTTTGCTGC CAAGATCTTC ACTAATGAAA GAAATCACCA GTGAGCTGCA CAGATTAGCC	360
	AAATACTGAG CTCATTAGAA CTACTAAGGC CTGGACATTG CTGCTTAATC CAGGACTCCT	420
30	GTAATTATCA GTCTTGCTT TGGAGCTTCC CATTGTGTAG CTGARAATTT GTCATATCTG	480
	CATTATAATC TAAGGCTCCA CATACTTAAT CCTGCTTCTC CCCCTTTTC TTTCCCTTTC	540
35	CCAGCGGTCA GCTCTGCTGC ATAGTCTGAA GACTTTCCCT GCCCAATCCT GATAAAATTC	600
	TTGCACTCGT AACCCCATCT CAGTGTCTG	629

40

## (2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:  
 45 (A) LENGTH: 1964 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

50	AAGAAGACAT GGAAATTGCT GAAGGATGTT TCAGGCATAT TAAGAAAATC TTTACGCAGC	60
	TTGAGGAATT CAGAGCCTCT GAATTGCTTC GAAGTGGACT GGACAGATCT AAATACCTTT	120
55	TAGTGAAGA AGCCAAAATT ATTGCTATGA CCTGTACTCA TGCTGCCCTA AAACGACATG	180
	ACTTGGTCAA GCTAGGTTTC AAGTATGACA ACATTTTGAT GGAAGAGGCT GCTCAGATTG	240
60	TGGAGATAGA AACTTTTATC CCTCTTCTTC TACAGAATCC TCAGGATGGA TTTAGCCGAC	300

	TAAAACGATG GATTATGATT GGCGATCATC ACCAGTTACC TCCAGTTATT AANGAACATG	360
	GCCTTCAAA AGTACTCAAA CATGGAGCAG TCTCTCTTCA CTCGCTTTGT TCGCGTTGGA	420
5	GTTCCGACTG TTGACCTTGA TGCTCAAGGG AGAGCCAGAG CAAGCTTGIG CAMCTNCTAC	480
	AACTGGCGAT ACAAGAATCT AGGAAACTTA CCCCATGTGC AGCTCTGCC AGAGTTTAGT	540
10	ACAGCAAATG CTGGCTTACT GTATGACTTC CAGCTCATTA ATGTTGAAGA TTTTCAAGGA	600
	GTGGGAGAAT CTGAACCTAA TCCCTTACTTC TATCAGAACATC TTGGAGAGGC AGAATATGTA	660
	GTAGCACTTT TTATGTACAT GTGTTTACTT GGTTACCCCTG CTGACAAAAT CAGTATTCTA	720
15	ACAACATATA ATGGCCAAAA GCATCTTATT CGCGACATCA TCAATAGACG ATGTGAAAC	780
	AATCCATTGA TTGGAAGACC AAACAAGGTG ACAACTGTTG ATAGATTTCA AGGTCAACAG	840
20	AATGACTATA TTCTTCTTTC TCTGGTACGA ACCAGGGCAG TGGCCATCT GAGGGATGTC	900
	CGTCGCTTGG TAGTGGCCAT GTCTAGAGCC AGACTTGGAC TTTATATCTT CGCCAGAGTA	960
	TCCCTCTTCC AAAACTGTTT TGAACTGACT CCAGCTTCA GTCAGCTCAC AGTCGCC	1020
25	CTTCATTGCA ATATAATTCC AACAGAACCT TTCCCAACTA CTAGAAAGAA TGGAGAGAGA	1080
	CCATCTCATG AAGTACAAT AATAAAAAT ATGCCCCAGA TGGCAAACCTT TGTATACAAC	1140
	ATGTACATGC ATTTGATACA GACTACACAT CATTATCATC AGACTTTATT ACAACTACCA	1200
30	CCTGCTATGG TAGAAGAGGG TGAGGAAGTT CAAAATCAAG AAACAGAATT GGAAACAGAA	1260
	GAAGAGGCCA TGACTGTTCA AGCTGACATC ATACCCAGTC CAACAGACAC CAGCTGCCGT	1320
35	CAAGAAACTC CAGCCTTCA AACTGACACC ACCCCCCAGTG AGACAGGAGC CACTTCACT	1380
	CCAGAAGCCA TCCCTGCTTT ATCTGAGACC ACCCCTACTG TGGTAGGAGC TGTATCTGCA	1440
	CCGGCAGAAG CTAACACACC TCAGGATGCC ACATCTGCC CAGAAGAGAC CAAGTAGCCA	1500
40	AACTGTAGTC CTTCTAAAGG AGGACATGCC AGTCAAAAG TCTGAGTAA GCTGTTTTT	1560
	GTATTTATA TTTGCTCTG CCATTTACT GTCACTAATT AATGTTAGT TCTTATATT	1620
45	GTTAACTGAT TTCGGTGTCT TGAATATATT TTTTAAATT ATGTGTATGA ACAATTCTAG	1680
	TTTCATTGAT TCAATCAGAA GAGCAAATAA CCATTCCCTT CATGTTTGA TCACTGAGTG	1740
	TGTCTGTAAT CATACTACA TTTAAATCAT TTTCTATGAA TATATAATAT ATACTTCACA	1800
50	TTTTTACTGTA ACTTCTCTAA AGAAGAGGAC AGAATATACT GGACTTAACC ACGAATACCC	1860
	TTGAGTGTCC AAATTGGAA GGAACTKGTT TCTTCYGTAA TACTAYCAA TGCTTAAATT	1920
55	CKGTTTCCCTT TTTCTTACCTT TTTGTTGCT GTCTTATGT AAAG	1964

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1522 base pairs  
 (B) TYPE: nucleic acid  
 5 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

10	CGTGTCCGCG CGCCTGGAG ACGCTGCCTC GGCCCGGACG CGCCCGGCC CCCGGGCTG	60
	GAGGTGGTC GCCACTGGGA CACTGTAAAC CAGGAGTRAG TCGGAGCTGC CGCGCTGCC	120
15	AGGCCATGGA CTGTGAGGTC AACAAACGGTT CCAGCCTCAG GGATGAGTGC ATCACAAACC	180
	TACTGGTGTG TTGCTTCCTC CAAAGCTGTT CTGACAACAG CTTCCGAGA GAGCTGGACG	240
	CACTGGGCCA CGAGCTGCCA GTGCTGGCTC CCCAGTGGGA GGGCTACGAT GAGCTGCAGA	300
20	CTGATGGCAA CGCGAGCAGC CACTCCCCTG TGGGAAGAAC AGAGGCAGAT TCTGAAAGTC	360
	AAGAAGACAT CATCCGGAAT ATTGCCAGGC ACCTCGCCA GGTCGGGAC AGCATGGACC	420
25	GTAGCATCCC TCCGGGCCTG GTGAACGGCC TGGCCCTGCA GCTCAGGAAC ACCAGCCGGT	480
	CGGAGGAGGA CGGAAACAGG GACCTGGCCA CTGCCCTGGA GCAGCTGCTG CAGGCCTACC	540
	CTAGAGACAT GGAGAAGGAG AAGACCATGC TGGTGCTGGC CCTGCTGCTG GCCAAGAAGG	600
30	TGCCAGTCA CACGCCGTCC TTGCTCCGTG ATGTCTTTCA CACAACAGTG AATTTTATTA	660
	ACCAGAACCT ACGCACCTAC GTGAGGAGCT TAGCCAGAAA TGGGATGGAC TGAACGGACA	720
35	GTTCCAGAAG TGTGACTGGC TAAAGCTCGA TGTGGTCACA GCTGTATAGC TGCTTCCAGT	780
	GTAGACGGAG CCCTGGCATG TCAACAGCGT TCCTAGAGAA GACAGGCTGG AAGATAGCTG	840
	TGACTTCTAT TTTAAAGACA ATGTTAAACT TATAACCCAC TTTAAAATAT CTACATTAAT	900
40	ATACTTGAAT GAAAATGTCC ATTACACGT ATTTGAATGG CCTTCATATC ATCCACACAT	960
	GAATCTGCAC ATCTGTAAT CTACACACGG TGCCTTTATT TCCACTGTGC AGGTTCCCAC	1020
45	TTAAAAAATTA AATTGAAAG CAGGTTCAA GGAAGTAGAA ACAAAATACA ATTTTTTTGG	1080
	TAAAAAAAAA TTACTGTTA TTAAAGTACA ACCATAGAGG ATGGTCTTAC AGCAGGCAGT	1140
	ATCCCTGTTG AGGAAAGCAA GAATCAGAGA AGGAACATAC CCCTTACAAA TGAAAATTG	1200
50	CACTCAAAAT AGGGACTATC YATCTTAATA CTAAGGAACC ACAATCTTC CTGTTAAAA	1260
	AACCACATGG CACAGAGATT CNGAACTAAA GTGCTGCACT CAAATGATGG GAAGTCCGG	1320
55	CCCCAGTACA CCAGGGCTT TGGACTTTT TCAACTTCGT TTCCCTTTGT TTGGANTCCA	1380
	AAAGAACACAC TTGTTGGTC TTAAAGGGT GTGAAGGTGA TTAAAGGGC CCAGGTCAGC	1440
	CACTGGTGTGG TTTACAAAAT CNGGGTAAC ACTGCATAC AACTTTTCC CNTTTCCATG	1500
60	NCATCAGGAC TTGCTAAAG AC	1522

## 5 (2) INFORMATION FOR SEQ ID NO: 42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 875 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

15	TGGGATTTCC CTTTATCATG GAGGCCTTGT CCCACTTCCT CTATGTCCT TTCCCTGGTG	60
	TCTGTGTC TGCGGCCATC TACACTGGCC TGTTCCCTCC TGAGACCAAA GGCAAGACCT	120
20	TCCAAGAGAT CTCCGAGGAA TTACACAGAC TCAACTTCCC CAGGCGGGCC CAGGGCCCCA	180
	CGTGGAGGAG CCTGGAGGTT ATCCAGTCAA CAGAACTCTA GTCCCAAAGG GGTGGCCGTA	240
	GCCAAAGCCA GCTACCGTCC TGTCCCTCTGC TTCCTGCCAG GGCCCTGGTC CTCAMTYCCT	300
25	YCTGCATTCC TCATTTAAGG AGTGTGTTATT GAGCACCCCT TGTGTGCGAGA CATGGCTCCA	360
	GGTGCTTAGC AATCAWIGGT GAGCGTGGTA TCCAGGCTAA AGGTAATTAA CTGACAGRAA	420
30	ATCAGTAACA ACATAATTAC AGGYTGGTTG TGGCAGYTCA TGACTGTAAT CCCAGCACTT	480
	TTGGGAGCCA AGGTGGGARG ATCAATTGAG GCCAGAGTTT GAAAMCAGCT AGGTAACATA	540
	GTGAGACCCC CTATCTCTAC AAAAAATTTC AACACATTAGC TGGGCATGGT GGTATGTGCT	600
35	AACAGCTCTA GCTACTCAGG AGGCTGAGGC AGCAGGATCA CTTGAGTCCA AGAGTTCAAG	660
	GTAGCAGTAA GCTACAATCA CACCACTGCA TGCCAGACTG GGTGACAGAG GGAGACTTCA	720
40	TCTCTTTAAA ACATAATAAT AAATAATTACA GACTCAGGAA ATGCAGTGAA AGAAAAATAC	780
	AGGTGGCCA GGTGAGGTGG CTGATGCCTG TAATCCCAGC ACTTTGGGAG GCCAAGATGG	840
	GAAGATIGCT TTGAGACCAG AAGTTTGAGA CCAGC	875

45

## (2) INFORMATION FOR SEQ ID NO: 43:

## 50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 843 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

	CCACGCGGT CCGNATCGTC CTTCCCTCAC TTCAGAGGGT GGCCAGAGCT GAATACCCAG	60
60	AGAGGGACAA GTAAGGGTCC AGTTCCAAAAT CATCATGAGG ATGTATCATC CCACGTGTCT	120

	CACCTGACAG TTACAGAGGA AACCCGCACC CAGAATGCAC GTGCTGTCTT ATGGGAACAC	180
5	TCAGGGCAGA GTGCTCAGGT CGGGCACAC TCGGGCTGTG CTTGGTCGTG CCATGGAATT	240
	CCTCAGGACT TTCTCAGCCT CCCTAACGGC AGAAGCCCCT TTACAGCAAG ACATTTACCG	300
	TTTGTCTGAA AATAGCCGAA CTGAGCTTTT CTTCAGGCTA TATGAGAAAGT CTCTAGACAG	360
10	TGGGCACCGT CAGAAAGCCC AGAGCCTTGT GATAGCTCCC ACCCTGCCTG GCTCAGATCT	420
	TCCCCATTTT TTTCTCTGG CACTAACCTC ACCTTTGTGTT TTTTGTTGTT TGTGTTGTT	480
15	TTTGTGTTG CAGAGTTGGA TTACAGAAAC TCCTATGAAA TTGAATATAT GGAGAAAATT	540
	GGCTCCCTCT TACCTGTAAG TTCGCTCGCC TCGGGCCACT TAGGGGACTC GCTTTCCCTGC	600
	CTTCAGGGGC CTCCCTCCCT GTGCAGAGTG TCTCTGGGAG CTCAGACCCC AAATCGAGTG	660
20	TTTCTGTGT ACACAGCTTC CGGGGTGCAC AGCAATGATG GACTGGGCT GGGGGGTTGA	720
	GGTTTGTACT CAATCCACTT CGTTTGACAT TTTCAGGGAG AAAATGATAG AATACAATTA	780
	GACGTCCCTGC AGAATTACTT TCCTAGACTG AGAAAGAGCT AGAGATTTCT TTAAAAAAA	840
25	AAA	843

30

(2) INFORMATION FOR SEQ ID NO: 44:

	(i) SEQUENCE CHARACTERISTICS:	
35	(A) LENGTH: 489 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:	
40	CTCTTAGGCT TTGAAGCATT TTTGCTGTG CTCCCTGATC TTCAAGGTAC CACCATGAAG	60
	TTCTTAGGAG TGCTGGTACT CTTGGGAGTT TCCATCTTTC TGGTCTCTGC CCAGAACCG	120
45	ACAACAGCTG CTCCAGCTGA CACGTATCCA GCTACTGGTC CTGCTGATGA TGAAGCCCCT	180
	GATGCTGAAA CCACTGCTGC TGCAACCACT GCGACCACTG CTGCTCCTAC CACTGCAACC	240
	ACCGCTGCTT CTACCACTGC TCGTAAAGAC ATTCCAGTTT TACCCAAATG GGTTGGGAT	300
50	CTCCCGAATG GTAGAGTGTG TCCCTGAGAT GGAATCAGCT TGAGTCTTCT GCAAATTGGTC	360
	ACAACTATTC ATGCTTCTTG TGATTTCATC CAACTACTTA CCTTGCTAC GATATCCCCT	420
55	TTATCTCTAA TCAGTTTATT TTCTTTCAA TAAAAAATAA CTATGAGCAA CAAAAAAA	480
	AAAAAAAAAA	489

60